SHAPE COMPLEXITY: RELATIVE AND EMERGENT, WITH APPLICATIONS
IN DEEP LEARNING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MAZLUM FERHAT ARSLAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

MARCH 2024

Approval of the thesis:

SHAPE COMPLEXITY: RELATIVE AND EMERGENT, WITH APPLICATIONS IN DEEP LEARNING

submitted by **MAZLUM FERHAT ARSLAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences**   ⸻

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering**   ⸻

Prof. Dr. Sibel Tarı
Supervisor, **Computer Engineering, METU**   ⸻

Examining Committee Members:

Prof. Dr. Halit Oğuztüzün
Computer Engineering, METU   ⸻

Prof. Dr. Sibel Tarı
Computer Engineering, METU   ⸻

Prof. Dr. Weihong Guo
Mathematics, Applied Mathematics, and Statistics, CWRU   ⸻

Assoc. Prof. Dr. Emre Akbaş
Computer Engineering, METU   ⸻

Assist. Prof. Dr. Venera Adanova
Computer Engineering, TED University   ⸻

Date:15.03.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Mazlum Ferhat Arslan

Signature   :

ABSTRACT

SHAPE COMPLEXITY: RELATIVE AND EMERGENT, WITH APPLICATIONS IN DEEP LEARNING

Arslan, Mazlum Ferhat

Ph.D., Department of Computer Engineering

Supervisor: Prof. Dr. Sibel Tarı

March 2024, 124 pages

Quantifying shape complexity is useful for a variety of applications including medical imaging and measuring the difficulty of samples in image datasets. However, the subject is underexplored, mostly due to its multifaceted nature. In this thesis, we start by proposing a benchmark dataset, subsets of which aim to account for the different aspects of the phenomenon. We compare a variety of shape complexity-related measures on the proposed dataset. Next, we propose a novel method that emphasizes the relative and emergent nature of shape complexity. The method operates in both continuous and discrete spaces of arbitrary dimensions. We demonstrate the properties of the method through extensive experiments and theoretical analysis. In the last part of the thesis, we turn to applications of the proposed measure. We obtain state-of-the-art results in domain generalization for prostate segmentation. As a separate application, we employ the proposed measure for curriculum learning on instance segmentation and image classification tasks on PASCAL VOC 2012 and CIFAR-10 datasets. We hypothesize the quantified shape complexity is an indicator of sample difficulties. Leveraging the estimated sample complexities, we devise curriculum strategies that lead to statistically significant performance increases on both tasks.

ÖZ

DERİN ÖĞRENMEDE UYGULAMALARIYLA GÖRELİ VE BELİRGEN ŞEKİL KARMAŞIKLIĞI

Arslan, Mazlum Ferhat
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Sibel Tarı

Mart 2024 , 124 sayfa

Şekil karmaşıklığının nicelenmesi tıbbi görüntüleme ve görsel veri kümelerindeki örneklerin zorluklarının ölçülmesi gibi uygulama alanlarında kullanım bulur. Ancak bu konudaki araştırmalar konunun çok yönlü yapısını kapsayıcı değildir. Bu tez kapsamındaki çalışmalarımız, şekil karmaşıklığının çeşitli yönlerini sınamayı hedefleyen bir değerlendirme veri kümesi önererek başlıyor. Önerilen veri kümesinde şekil karmaşıklığı ölçen metotları karşılaştırıyoruz. Çalışmalarımız kapsamında şekil karmaşıklığının göreceli ve belirgen olma özelliklerini hedefleyen yeni bir yöntem öneriyoruz. Bu yöntem çok boyutlu sürekli ve süreksiz uzaylarda geçerlidir. Önerilen yöntemin özelliklerini kapsamlı deneyler ve kuramsal incelemeler aracılığıyla gösteriyoruz. Tezin son bölümünde, önerilen ölçümün uygulamalarını ele alıyoruz. Prostat bölütlemesi için alan genelleştirmesi uygulamalarında alandaki en iyi sonuçları elde ediyoruz. Ayrı bir uygulama olarak, önerilen ölçümü PASCAL VOC 2012 ve CIFAR-10 veri kümelerinde örnek bölütleme ve görüntü sınıflandırma görevlerinde müfredat öğrenme için kullanıyoruz. Nicelenen şekil karmaşıklığının örnek zorluklarının bir göstergesi olduğu varsayımına dayanarak tahmin edilen karmaşıklıkları

geliştirdiğimiz müfredat yöntemleriyle kullanıyoruz. Bu yaklaşımla, her iki görevde de istatistiksel olarak anlamlı performans artışları elde ediyoruz.

Anahtar Kelimeler: şekil karmaşıklığı, morfolojik diferansiyel denklemler, tıbbi görüntü bölütleme, alan genelleştirme, müfredatla öğrenme

To my family,

with all the suns

ACKNOWLEDGMENTS

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

FIGURES

xviii

CHAPTER 1

INTRODUCTION

Science seeks explanation to complex phenomena, making complexity itself a subject of scientific interest. By this nature, complexity studies span various disciplines. However, despite our ability to perceive complexity, its formal definition remains elusive ([1, 2, 3], or more recently [4]). Thus, the *modus operandi* for complexity studies are discipline dependent. For example, in computer science the scaling of optimal algorithms with input size is a relevant inquiry, whereas for biology whether complexity increases during evolution, or, in case of dynamical system studies, understanding chaos, are of interest.

The generic nature of complexity studies allows for the exploration of a wide range of objects, including shapes, problems (*e.g.* sorting), complex wholes (*e.g.* biological organisms), and processes (*e.g.* dynamics of a double pendulum). In this thesis, we focus on shapes.

1.1 Motivation and Research Questions

The exploration of complexity in shapes is desirable for several reasons. Firstly, shapes serve as a bridge between the abstract and applied subjects since they can be situated between the formal and the physical. Compared to the formal (canonically, bit strings[1]) shapes are semantically richer: they come equipped with a sense of topology allowing for multi-dimensional analysis, and their discrete and continuous representations have clear correspondence. Compared to the physical they are

---

[1] It can be argued that bit strings are one-dimensional shapes with multiple components. However, we consider the topological structure in higher dimensions non-trivial, thus, make a distinction between the two.

more manageable: they are not temporal, and for digital shapes, they are finitely representable. Secondly, owing to their geometric nature, shapes and their relations are intuitively interpretable. Given the exploratory nature of the current study, we find this valuable. Last but not least, shape complexity is useful on its own. It finds applications in clinical neuroscience [5], analysis of neural networks in terms of decision hypersurfaces [6], and data encryption [7].

Despite the existence of a variety of shape complexity measures, to the best of our knowledge, a comparison of the methods is missing in the literature. However, such comparisons help the accumulation of methodological improvements and findings in the field. This leads us to our first research question:

1. How do different shape complexity measures compare to each other in terms of different aspects of complexity?

We address this question in Chapter 2.

"The whole is something beside the parts" [8]. Similarly, the building blocks of a complex whole do not explicitly manifest the complexity of the whole. Rather, complexity emerges from interactions among the building blocks. Therefore, explanations of complexity in terms of only local quantities fail to capture it, though, correlations may exist. Seeking a measure of shape complexity that considers both local and global interaction of shape parts and that encodes each part's contribution towards the overall complexity, we pose the following question:

2. How can the emergent aspect of complexity be addressed to yield a feasible and consistent shape complexity encoding?

Here, the feasibility is in the sense that the measure quantifies shapes that are implicitly assumed as simplest as indeed so. Consistency is with respect to the perceptual expectations under several controlled deformations of arbitrary shapes. This question is addressed in Chapter 3 and Appendix A.

As we will argue in detail in Chapter 4, complexity is relative. Therefore, we address:

3. How can we incorporate the relative aspect of complexity in addition to the concerns in the previous research question?

Our final research question considers the practical applications of the proposed shape complexity measure:

4. How can shape complexity be used in machine learning applications?

Chapter 5 and Chapter 6 report our efforts in answering this question.

1.2 Scope of the Thesis

In this thesis our focus is on the quantifications of shape complexity, addressing its relative and emergent aspects, and its applications.

For quantifications of shape complexity, the proposed benchmark dataset is comprised of three collections that measures the correlations of methods with the noise present in shapes, expert-rankings of design object, and segmentations of 3D shapes. We provide ground truth labels for each. Further, we compare both 2D and 3D methods which include shape complexity and other related measures, such as those quantifying convexity or circularity. Our evaluations are in terms of pairwise agreement with ground truth.

In our own considerations, shapes are bounded, connected, binary objects of arbitrary dimensionality. Complexity is taken to be a measurable quantity indicating the deviation of an object from *simple*, where simple is an object, defined implicitly or explicitly. The relative aspect is accounted for when simple is an explicit parameter. The emergent aspect is addressed by using both local and global shape features through a partial or morphological differential equation formulation.

Our focus is on structural complexity [9]; thus, a time-like dimension of objects is left out. Further, the existence of an underlying space-like structure in which the object information is carried is assumed. Nonetheless, we try to keep our framework general enough so that extensions to more general objects (such as graphs) are possible with minimal effort in future work. Also, since we are after a measure of complexity, computability is of concern. This leads the work to be more focused on objects in computational grids, *i.e.*, the digital shapes.

Complexity measures can be broadly classified into three categories [10]: those related to difficulty of description, difficulty of creation, and degree of organization. The proposed complexity measure relates the most to the last category. Notably, Gell-Mann [2] considers measures of 'something like complexity' and not of 'complexity', stressing that a single measure would be incapable of capturing all aspects of complexity. Inevitably, in an attempt to explore certain aspects, our considerations result in a measure of something like complexity as well.

We attempt to establish the relevance of our method to complexity by both quantitative and qualitative analyses on two and three dimensional shapes (Chapter 2, 3, 4, and Appendix A), and also, by means of medical imaging and computer vision applications that serve as proxy tasks relating to current scientific interests (Chapter 5 and 6). In medical imaging, we consider the task of single domain generalization for 3D prostate MRI segmentation with a UNet-like segmentation backbone. We show that our measure can be effectively integrated to increase performance in terms of Dice score compared to the baseline and two state-of-the-art methods.

For computer vision applications, we consider instance segmentation and image classification tasks. We incorporate our measure via curriculum learning. Compared with the baseline method, curriculum strategies based on the sample complexities acquired by our method yield statistically significant performance increases.

1.3 Contributions and Novelties

Our contributions are as follows:

- We propose a benchmark dataset for shape complexity and compare 18 two and three dimensional methods on three different tasks.

- We propose a semantically rich relative shape complexity measure. The framework for the proposed measure is valid in both continuous and discrete, arbitrary dimensional spaces.

- We propose a novel application of shape complexity to medical image segmentation using the proposed relative shape complexity measure. We achieve

state-of-the art results on the single source domain generalization setting by emphasizing the learning on the low complexity regions.

- We further demonstrate the relevance of the proposed measure to complexity by means of estimating the sample image difficulties which are then used for curriculum learning in two different tasks, instance segmentation and classification.

The outcomes of the thesis are materialized in the following publications:

- M. F. Arslan, A. Haridis, P. L. Rosin, S. Tari, C. Brassey, J. D. Gardiner, A. Genctav, and M. Genctav, "SHREC'21: Quantifying Shape Complexity," Computers & Graphics, vol. 102, pp. 144–153, 2022,

- M. F. Arslan and S. Tari, "Local Culprits of Shape Complexity," in International Conference on Scale Space and Variational Methods in Computer Vision, pp. 91–99, Springer, 2021.

The following paper is in preparation:

- M. F. Arslan and S. Tari, "Measuring relative shape complexity," (in preparation)

1.4   The Outline of the Thesis

In Chapter 2, we start with a review of the literature on shape complexity. We propose a benchmark dataset that is composed of three separate collections. Each collection aims to account for a different aspect of shape complexity. A comparison of 18 methods of shape complexity on this benchmark is given.

Chapter 3, together with the associated Appendix A, aims at addressing the emergent aspect of shape complexity. We propose a measure that is formulated for continuous domains of arbitrary dimensions. The properties of the proposed measure is discussed in detail via illustrative examples.

A measure of shape complexity, accounting for the phenomenon's relative nature, is presented in Chapter 4. We start with a discussion on the relativity of complexity and proceed with defining the notion of simplicity with respect to an arbitrary prototype shape, discussing its relevance to compressibility. Our considerations are valid for both continuous and discrete domains. The chapter is concluded by presenting proof-of-concept results.

In Chapter 5, we propose a novel application of the relative shape complexity measure for domain generalization in medical image segmentation. The application showcases the use of the proposed measure in the three-dimensional setting and also supports the claim that the complexity-encoding field is semantically rich. The proposed method achieves state-of-the-art results without introducing additional parameters to the segmentation backbone and with negligible computational cost during training.

To further demonstrate the relevance of the proposed measure's relation to complexity and deep learning applications, in Chapter 6 we consider curriculum learning on two computer vision tasks. After quantifying the shape complexity associated with the images in the PASCAL VOC 2012 and CIFAR-10 datasets, we experiment with two network architectures, Mask R-CNN for the former for instance segmentation and a modified ResNet-18 for the latter for image classification, using curricula based on the quantified sample complexities. In each case, we observe significant improvements in the in-domain generalization performance of the networks.

Finally, to ease the readability of the thesis, Chapter 7 gathers the mathematical and numerical considerations related to the proposed measures. These include analytical solutions to the considered nonlinear partial differential equations, some mathematical properties of the solutions, numerical schemes, and fast solvers using multigrid methods for constructing the scalar fields.

CHAPTER 2

SHAPE COMPLEXITY AND A PROPOSED BENCHMARK

This chapter is based on our paper [11]. In order to maintain the integrity of the thesis, modifications are made to the content of the published paper.

The main contributions of the work are as follows:

- The creation of two novel shape collections, each accompanied by its ground truth (GT). Additionally, a previous segmentation benchmark has been repurposed for evaluating complexity measures.

- A systematic evaluation of the performance of various two-dimensional (2D) and three-dimensional (3D) classical and recent shape complexity measures.

- Assessment of the similarities and differences among various measures through pairwise correlations and clustering. This analysis is based on their performance with respect to multiple GTs.

Figure 2.1 presents a graphical illustration of the overview of the proposed benchmark.

2.1 Related Work on Shape Complexity

Shape complexity has been studied in various fields including psychology [12, 13, 14], computer vision [15, 16], mathematical morphology [17], and design [18, 19]. It finds applications in shape retrieval [20, 21], measuring neurological development [22, 23], and computer-aided design (CAD) [24, 25, 26]. The perspectives on shape
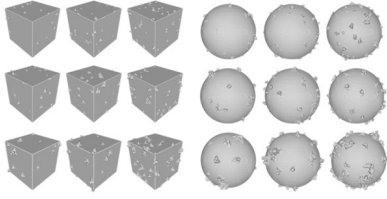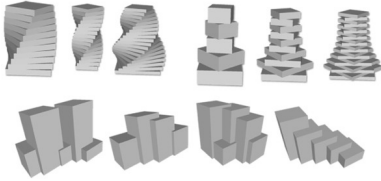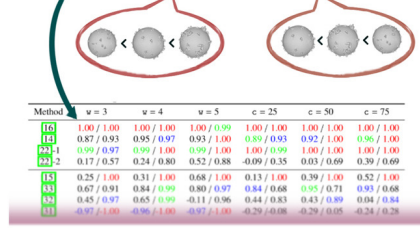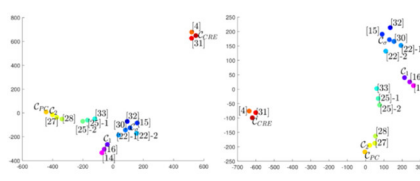
Figure 2.1: Graphical abstract for the proposed benchmark dataset and methodology.

complexity differ dramatically between disciplines due to their respective methodologies and objectives. For example, works in psychology rely on human experiments while in computer vision, quantification of Euclidean notions such as curvature are of interest.

Early work on shape complexity appears in the literature of experimental psychology. One of the first measures of complexity for polygonal shapes can be found in [27]. Attneave [12] conducted human experiments to seek correlations of shape complexity with scale, curvedness, symmetry and number of turns. Due to diverse responses from human subjects, Attneave concluded that shape complexity is not well-defined. Considering circles as the simplest shapes, P^2/A emerged as a natural measure for shape complexity and was used in several works [12, 18], either as a direct measure or alongside other indicators. In most other works [28, 17, 29, 30, 31, 32], tools from information theory applied on various geometric features are used to quantify shape complexity. Rossignac [33] provides a classification of shape complexity that focuses on measuring different aspects of computer representations for 3D shapes.

The variety of approaches to quantifying shape complexity further supports the claim that complexity can have various meanings depending on the chosen approach in a particular research area and for the specific task at hand.

Despite the wide range of perspectives and applications, a pattern can be identified in the works on shape complexity: shape complexity is related to the local and/or global regularity of some feature of the shape. For example, Page *et al.* [28], Chen and Sundaram [16], and Matsumoto *et al.* [34] considered curvature, Gardiner *et al.* [35] and Brinkhoff *et al.* [36] considered convexity, and Saleem *et al.* [15] considered view similarity as an indicator of shape complexity. In [32, 37], emergent features of rectangularity are used to quantify shape complexity. For measuring regularity, Shannon entropy, correlates of Kolmogorov complexity, or the mean of the feature distribution are common options.

The works [32, 31, 30] construct scalar fields over a shape via screened-Poisson-like differential equations, each with distinct approaches. In [32], the authors consider

$$\left(\Delta_\infty - \frac{1}{\rho^2} \right) f = -1 \quad \text{subject to } f\Big|_{\partial S} = 0$$

for a shape S and quantify the level set overlap between the solution f and the distance transform of S. Here, Δ_∞ is the infinity Laplacian which is the minimizer of $\int |\nabla f|^p$ as $p \to \infty$. The targeted level sets determine the trade-off between boundary details and global shape properties. This allows for multi-scale measurements of shape complexity.

In Discrepancy, [31], the equation

$$\left(\Delta - \frac{1}{A^2} \right) f = 0 \quad \text{subject to } f\Big|_{\partial S} = 1$$

is deployed as a measure of local circularity in 2D, where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. In order to measure shape complexity, the authors use entropy on the probability density functions (PDF) acquired from the constructed field's histogram.

In PARCELLIN distance, [30], a family of equations are considered

$$(\Delta - \alpha) \, \Phi_i = -f_i$$

where f_i are the source functions with $i \in [1, 2, \cdots, N]$. The source functions stand for separating the volume bounded by the shape boundary into central and outer regions, and vary smoothly with the level sets of the distance transform of the shape. After acquiring the fields $\{\Phi_i\}_{i=1}^{N}$, the fields are aggregated by assigning each shape location the number of solutions in which the location falls into the outer region, *i.e.*,

attains a negative value. The entropy of the PDF arising from the aggregated field near shape boundary is used for measuring the shape complexity.

In [36], a heuristic measure of shape complexity is defined for polygons as a combination of frequency and amplitude of the boundary vibration, and deviation of the shape from its convex hull. The suitability of the approach is shown on maps of geographical objects. By using a discrete approximation for the curvature, [28] measure the shape complexity as the entropy of the PDF derived from the histogram of the approximated curvature. We also define an alternative version that uses cumulative residual entropy [38] $\mathcal{C}_{CRE}$ instead of Shannon entropy. More recently, a similar measure of complexity that leverages curvature is proposed by [34].

Fractal dimension [39] of shape boundaries are used in [40] to quantify the regularity of shapes. In their approach, the scale parameter determines the trade-off between local and global information. Two other methods, [41] and [42], that use fractal dimensions are also treated as shape complexity measures.

In addition to fractal dimensions, measures of convexity can serve as indicators of shape complexity. Here, we consider some convexity measures. Two standard convexity measures involve the ratio of areas $\mathcal{C}_1$ and ℓ_2 perimeters $\mathcal{C}_\in$ of a shape S with its convex hull $[S]$,

$$\mathcal{C}_1 = \frac{\mathrm{Area}(S)}{\mathrm{Area}([S])}, \qquad \mathcal{C}_2 = \frac{|\,\partial[S]\,|_2}{|\partial S|_2}$$

$|\partial S|_p$ denotes the ℓ_p perimeter length of shape S. A more involved measure of convexity, [43], applies polygonal convexification by flipping the concavities of a shape about their corresponding edges. An alternative scheme can be defined by applying *flipturns*, *i.e.*, reversing the order of vertices while applying a flip to the concavity. The processes result in a convex polygon in a finite number of steps. Two measurements of convexity are then defined as the ratio of the areas of the original and convexified polygon for the two scenarios, which are denote as [43]-1, and [43]-2 in the following. Another measure that relates to convexity is that of [44], given by

$$\mathcal{C}(S) = \min_{\theta \in [0,2\pi]} \frac{|\partial \mathcal{R}(S_\theta)|_2}{|\partial S_\theta|_1}$$

10

where S_θ refers to the shape rotated by angle θ, and $\mathcal{R}$ is the minimum area bounding rectangle of its parameter.

In [35], alpha-shapes [45] with N different α parameters are acquired to account for the shape at various scales. For each α, the ratios of the volumes of the original shape and the alpha-shape are obtained, which represent the shape in $\mathbb{R}^N$. Finally, the PCA scores corresponding to the first two principal components are used as indicators of shape complexity.

2.2 Proposed Benchmark Dataset

The lack of benchmark datasets for shape complexity poses a significant challenge in the current state of the field. This causes works in the literature to opt for suboptimal analyses, relying only on visual results. Thus, it is our aim in proposing a benchmark dataset to address and explore various aspects of shape complexity which can help researchers to test and compare their methods against others in the literature.

Due to the lack of a canonical definition for shape complexity, expecting the methods to correctly guess pre-defined linear relationships between shapes may not make sense. Thus, we propose to explore complexity using multiple tasks and multiple shape collections.

The proposed benchmark dataset is composed of three collections. The first collection is composed of shapes obtained by introducing additive or subtractive noise to two basic shapes: sphere and cube. The purpose is to investigate the relation of complexity to the noise level. The second collection is composed of artificial three-dimensional shapes constructed by transforming and combining multiple elements, and evaluated by experts to provide GT. The purpose of this collection is to investigate the shape complexity measures in relation to perceptual categories. The final collection is an existing three-dimensional shape dataset which was originally developed as a segmentation benchmark. We repurpose this data and use the segmentation GT as a means to investigate 3D shape complexity via a proxy task.

Each collection in the dataset is a set of three-dimensional objects with different characteristics. Similarly, the GTs for each collections are defined and obtained through

distinct methods. In this way, we aim to account for the different aspects of shape complexity.

2.2.1 Dataset

The proposed benchmark dataset consists of three collections each aiming to account for a different aspect of shape complexity. The first two collections are created synthetically, whereas the third is an existing collection consisting of natural shapes. The GT for the first collection is based on the parameters used in creating the collection. For the second collection, the GT is provided by two design experts on the final design object. The purpose of the third collection is to investigate how estimated complexity is related to the number of parts perceived by humans, which we hypothesize is related to shape complexity.

2.2.1.1 Collection 1 – Perturbed basic shapes

In this collection, we aim to explore the correlation between shape complexity and noise. Specifically, we consider two base shapes, cube and sphere, and introduce perturbations to each shape with three control parameters, explained in detail in the following.

A cube of side length 199 voxels and a sphere of radius 100 voxels are stochastically perturbed additively and subtractively, separately. This forms four families (additively/subtractively perturbed cubes/spheres). The algorithm used in adding noise to a shape introduces a perturbation at a random location on the shape's boundary at each application. The algorithm has two parameters: *i)* width (w) determining the area of effect of the perturbations and *ii)* number of times of application (c) determining how many times a local perturbation is introduced. Both parameters are set to three different values, $w \in \{3,\ 4,\ 5\}$ and $c \in \{25,\ 50,\ 75\}$. This results in a group of nine shapes. A sample group for an additively perturbed cube is displayed in Figure 2.2. Fifty such groups form a family.
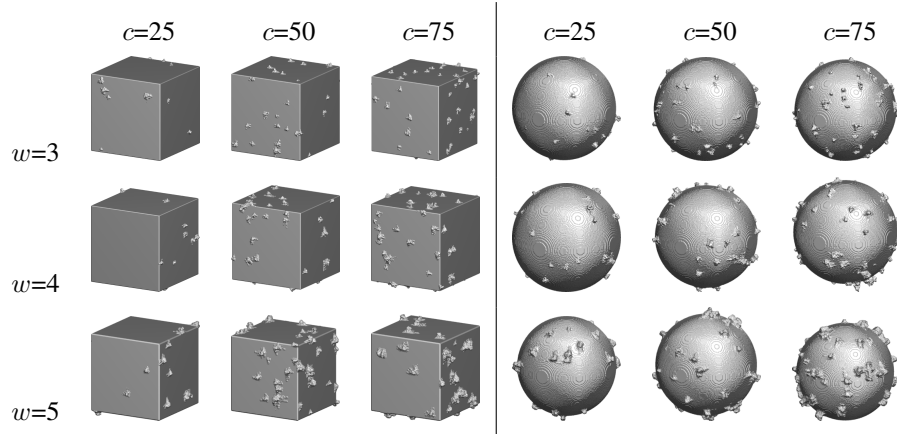
Figure 2.2: Sample groups of additively perturbed cubes and spheres.

2.2.1.2 Collection 2

Collection 2 consist of fifty shapes, equally divided into two families. The shapes are designed in a way to resemble abstract architectural models. The building blocks are cuboids that form a connected aggregate for each shape.

In the first family, cuboids are stacked to have a constant total height. The underlying parameters are the number, height and rotation of the cuboids. In Figure 2.3, samples from the first family are shown (the colors are for the purposes of visualization only).
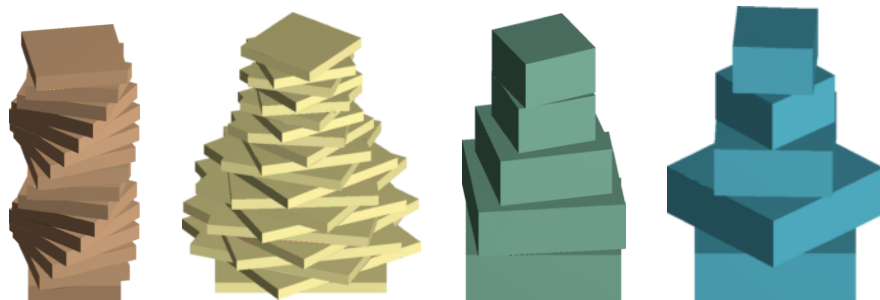
Figure 2.3: Sample shapes from the first family of Collection 2.

The second family is generated by varying the spatial organization of a number of cuboids. The number of cuboids ($n \in \{3, 4, 5\}$), their relative locations, and the height of each cuboid are the used parameters. The cuboids are allowed to have overlapping regions, as exemplified in Figure 2.4.
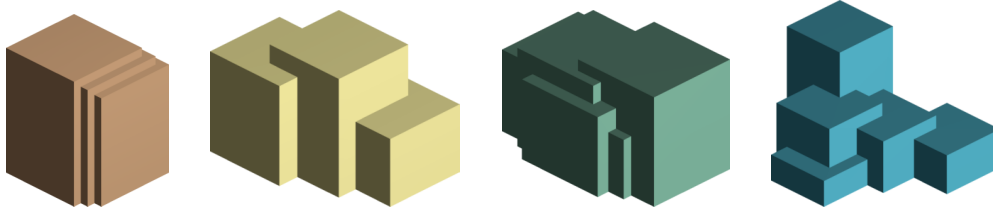
Figure 2.4: Sample shapes from the second family of Collection 2.

2.2.1.3 Collection 3

Collection 3 is composed of the shapes in Princeton mesh segmentation benchmark dataset [46]. We use this set with the primary objective of exploring how shape complexity measures correlate with the uniformity of the number of segments of the segmentations of the shape. The benchmark consists of 380 shapes across 19 categories and their human-generated segmentations. As opposed to the synthetic shapes in the first two collections, the shapes in the benchmark are natural. As such, they have a particular semantic content, which may affect the perception of complexity. The availability of manual segmentations for this collection makes it a fitting candidate for exploring complexity by using segmentation as a proxy task.

2.2.1.4 2D Collections

Most of the shape quantifying methods in the literature work exclusively in 2D. To include such methods into this study, we create the 2D analogues of the shapes. We generate twelve 2D silhouettes of each shape in the above collections from the views determined by the azimuthal angles ($\{0°, 30°, 60°, \ldots 330°\}$) and the elevation angle ($30°$). The resulting silhouettes of a shape are kept similar in size so that the collections do not pose a challenge in terms of scale-invariance. The compared 2D methods report the averaged score over the twelve silhouettes as the measure of complexity for the corresponding 3D shape.

We exclude the families consisting of subtractively perturbed spheres and cubes because the resulting silhouettes highly depend on whether the perturbations appear on the 2D boundary of a given view or not, rather than the controlling parameters.

2.2.2 Ground truths and evaluation methods

Collection 1. For the first collection, the two parameters w and c used in creating the shapes constitute the GT. We expect the complexity scores to increase as either of the parameters increase. The performance of the methods are measured in a controlled experiment manner: we keep one of the parameters fixed and let the other vary. The performance of a method is then measured by averaging the Kendall rank correlation coefficient over the groups. This results in six measures of performance (one for each value of the parameters) for a family.

Collection 2. In Collection 2, the GT is provided manually by two experts on CAD and 3D shape modeling.

For the first family, the evaluation produced the following five groups in which shapes are listed in increasing order of complexity:

Group 1: $(16, 14, 12, 17, 18, 19, 20)$

Group 2: $(15, 13, 24, 22)$

Group 3: $(23, 21, 25)$

Group 4: $(11, 10, 7, 1, 2)$

Group 5: $(9, 8, 4, 5, 3)$

For the second family, the evaluation produced the following six groups in which shapes are considered to be of equal complexity:

Group 1: $(18, 24, 17, 20)$

Group 2: $(25, 21, 22, 19)$

Group 3: $(5, 3, 2)$

Group 4: $(23, 16, 15)$

Group 5: $(6, 4, 14, 12, 13)$

Group 6: $(7, 9, 10, 1, 11, 8)$

As the shapes from different groups are incommensurate, GT is provided only for the shapes in the same group. We measure the performance of the methods using the

Kendall rank correlation coefficient since we have a total order on the groups of the first family. For the second collection, we measure the uniformity of in-group scores. The scores are first normalized to the range $[0, 1]$. The pairwise absolute differences of the normalized scores are summed to yield the performance measure of a group. Note that scores closer to zero indicate a better performance, in contrast to the rest of the performance measures.

Collection 3. The shapes in the third collection are segmented by both humans and computer algorithms in [46]. We interpret the data collected from humans as an indicator of a shape's complexity. For each shape, there are 11 human-generated segmentations and 7.9 segments, on average. The fact that the human annotations differ is consistent with the ill-posed natures of both complexity and the segmentation task.

For this collection, we use two GT: one is the mean (μ) and the other is the standard deviation (σ) of the number of segments. For each GT, we calculate *i)* Kendall rank correlation coefficient over all the shapes in the collection which we refer to as $\tau_{\mu_{\text{all}}}$ and $\tau_{\sigma_{\text{all}}}$ in Table 2.4 *ii)* the averaged coefficients $\frac{1}{N} \sum_i \tau_i$ where τ_i is the correlation coefficient for the ith category, referred to as $\tau_{\mu_{\text{cat}}}$ and $\tau_{\sigma_{\text{cat}}}$.

2.3 Results

Since our GTs provide the order information and not complexity quantifications, we compare the methods in terms of the orderings they suggest. Hence, we use only Kendall rank correlation as a robust rank correlation measure. We report Kendall rank correlation coefficients between the participating methods and the GTs in Tables 2.1-2.4. In the tables, we mark the scores of the best performing methods with red, the second best performers with green, and the third best performers with blue.

2.3.1 Collection 1

Additive perturbations. The Kendall rank correlation coefficient (τ) for the additively perturbed cubes and spheres are given in Table 2.1. The method proposed by

[32] effectively orders all pairs for cubes, *i.e.* $\tau = 1.00$, while the method by [31] achieves the same for spheres. In both cases, [30] and [35]-1 follow very closely.

Some of the methods ([28], [40], $\mathcal{C}_{CRE}$, [44], [43]-1, [43]-2 and $\mathcal{C}_2$) achieve strong correlations with the GT when the parameter c is varied, yet the correlations weaken when the varying parameter is w. This suggests that it is easier to account for the number of perturbations than it is for the magnitude of perturbations.

Additionally, we observe that the performance of some methods varies significantly when the base shape is changed, particularly for [31], [41], [42] and $\mathcal{C}_1$. Notably, all of these methods achieve better scores on the spheres.

Comparing 3D methods with 2D ones, we see that 3D methods [32], [30] and [35]-1 consistently perform well whereas the best performing 2D methods are inconsistent. For example, [31] only achieves high scores for the sphere-related tasks and [36], [28], $\mathcal{C}_{CRE}$, $\mathcal{C}_{PC}$, and so on, attain high scores only when the width parameter w is kept fixed.

Subtractive perturbations. Four methods have been tested on the subtractively perturbed cubes and spheres. The performances of these are reported in Table 2.2. Three of the methods perform consistently well in all cases. For the cubes, [30] ranks the first in all measurements, and for the spheres, there is no clear winner.

In the last rows of Tables 2.1 & 2.2 we provide the mean of the absolute scores, denoted as MA. The mean absolute scores show that the most challenging case is the cubes with $c = 25$ for both perturbation types. We also note that, for additive perturbations, it appears significantly harder for the considered methods to correlate with the GT when the parameter w is varied compared with the case when c is the varying parameter.

2.3.2 Collection 2

The results for both families of Collection 2 are presented in Table 2.3. For the first family of shapes in the collection, higher absolute values indicate better performance. Conversely, for the second family, values near zero are preferred.

Table 2.1: The averaged Kendall τ for the additively perturbed cubes (the first value) and spheres (the second value).

Method	$w = 3$	$w = 4$	$w = 5$	$c = 25$	$c = 50$	$c = 75$
[32]	1.00 / 1.00	1.00 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
[30]	0.87 / 0.93	0.95 / 0.97	0.93 / 1.00	0.89 / 0.93	0.92 / 1.00	0.96 / 1.00
[35]-1	0.99 / 0.97	0.99 / 1.00	0.99 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00
[35]-2	0.17 / 0.57	0.24 / 0.80	0.52 / 0.88	-0.09 / 0.35	0.03 / 0.69	0.39 / 0.69
[31]	0.25 / 1.00	0.31 / 1.00	0.68 / 1.00	0.13 / 1.00	0.39 / 1.00	0.52 / 1.00
[42]	0.67 / 0.91	0.84 / 0.99	0.80 / 0.97	0.84 / 0.68	0.95 / 0.71	0.93 / 0.68
[41]	0.45 / 0.97	0.65 / 0.99	-0.11 / 0.96	0.44 / 0.83	0.43 / 0.89	0.04 / 0.84
[40]	-0.97 /-1.00	-0.96 /-1.00	-0.97 /-1.00	-0.29 /-0.08	-0.29 / 0.05	-0.24 / 0.28
$\mathcal{C}_{CRE}$	0.97 / 1.00	0.95 / 1.00	0.96 / 1.00	0.33 / 0.12	0.47 /-0.07	0.47 /-0.35
[28]	0.93 / 1.00	0.96 / 1.00	0.96 / 1.00	0.20 / 0.28	0.39 / 0.17	0.32 /-0.21
[34]	0.79 / 0.87	0.87 / 0.93	0.87 / 0.93	0.83 / 0.83	0.88 / 0.83	0.88 / 0.79
$\mathcal{C}_{\sigma}$	0.51 / 0.72	0.60 / 0.80	0.64 / 0.87	0.68 / 0.60	0.68 / 0.63	0.69 / 0.64
$\mathcal{C}_1$	-0.73 /-0.97	-0.83 /-0.97	-0.76 /-0.95	-0.77 /-0.89	-0.81 /-0.95	-0.93 /-0.96
$\mathcal{C}_{PC}$	0.93 / 0.96	0.91 / 0.93	0.88 / 0.97	0.61 / 0.53	0.71 / 0.51	0.64 / 0.36
[36]	0.93 / 1.00	0.95 / 0.99	0.91 / 0.99	0.73 / 0.69	0.83 / 0.77	0.79 / 0.64
[44]	-1.00 /-0.99	-0.96 /-0.99	-0.91 /-0.96	-0.68 /-0.59	-0.77 /-0.68	-0.76 /-0.61
$\mathcal{C}_2$	-0.95 /-0.99	-0.96 /-0.99	-0.93 /-0.99	-0.67 /-0.53	-0.80 /-0.63	-0.77 /-0.51
[43]-1	-0.96 /-0.99	-0.93 /-0.99	-0.92 /-0.99	-0.59 /-0.68	-0.76 /-0.76	-0.76 /-0.64
[43]-2	-0.96 /-0.99	-0.95 /-0.99	-0.92 /-0.99	-0.65 /-0.68	-0.76 /-0.76	-0.75 /-0.63
MA	0.79 / 0.94	0.83 / 0.96	0.82 / 0.97	0.60 / 0.65	0.68 / 0.69	0.68 / 0.68

The first family of shapes. For the first family of Collection 2, based on the summed scores, the top five methods are $\mathcal{C}_{CRE}$, $\mathcal{C}_{PC}$, [40], [28] and $\mathcal{C}_2$. Note that all of these are 2D methods, three of which are convexity measures. The best performing 3D method is [30] and places the sixth. [34] and [35]-2 perform the poorest on this family both having almost no correlations with the GT considering all of the groups.

For each group except the fifth, there is at least one method that completely agrees (or disagrees) with GT. For Group 5, none of the considered methods is able to capture the notion of complexity that exactly agrees with GT. The MA scores are in alignment with this, indicating that Group 5 ($MA = 0.49$) is indeed the most challenging group.

Lastly, we observe that some of the methods ([41], [42], [31] and [35]-2) have both strongly positive and strongly negative correlations with GT.

The second family of shapes. For the second family of Collection 2, we start by

18

Table 2.2: The averaged Kendall τ for the subtractively perturbed cubes (the first value) and spheres (the second value).

Method	$w = 3$	$w = 4$	$w = 5$	$c = 25$	$c = 50$	$c = 75$
[32]	1.00 / 0.99	1.00 / 0.99	1.00 / 1.00	0.67 / 0.92	0.89 / 1.00	0.91 / 0.97
[30]	1.00 / 0.89	1.00 / 0.88	1.00 / 0.93	0.97 / 0.97	0.97 / 1.00	0.99 / 1.00
[35]-1	0.68 / 0.81	0.81 / 0.95	1.00 / 1.00	0.77 / 0.83	0.91 / 1.00	0.97 / 1.00
[35]-2	0.21 / 0.48	0.48 / 0.83	0.87 / 0.96	0.37 / 0.61	0.69 / 0.80	0.81 / 0.95
MA	0.72 / 0.79	0.82 / 0.91	0.97 / 0.97	0.70 / 0.83	0.87 / 0.95	0.92 / 0.98

remarking that the reported scores indicate better performances when they are close to 0, in contrast with the other reported scores. Similar to the case in the first family, 2D methods take the lead (listed from best to worst: [43]-1, [36], $\mathcal{C}_2$, $\mathcal{C}_1$, [34]), with the best performing 3D method ([35]-1) placing the 8th. The worst performing method is [28]. This is interesting because it is also the third best performing method in the first family. In a similar manner, we note that there is no overlap between the top five performers of the two families except for $\mathcal{C}_2$. This indicates that the two families indeed account for different aspects of shape complexity.

Based on the overall scores, the top three performers for both families are 2D methods. The highest scoring 3D method for the first family is [30], and [35]-1 for the second family.

2.3.3 Collection 3

The Kendall rank correlation coefficients computed for Collection 3 are reported in Table 2.4. First, we note that, even for the best-performing methods, all correlations exhibit considerably weaker values when compared to the scores obtained from the other two collections. The best performers are [35]-2, [34], [28] and [28] for $\tau_{\mu_{\text{cat}}}$, $\tau_{\mu_{\text{all}}}$, $\tau_{\sigma_{\text{cat}}}$ and $\tau_{\sigma_{\text{all}}}$, respectively. We observe that all of the methods, except [35]-1, perform better when the correlations are computed over the whole collection, regardless of GT.

For the tasks of this collection we observe that [35]-2 outperforms [35]-1. This is interesting because [35]-1 is a better performer for the majority of tasks on the other two

Table 2.3: The Kendall τ for the first family and the non-uniformity measurements for the second family of Collection 2 reported as the first and second scores, respectively.

Method	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Sum
[32]	0.33 / 2.91	0.67 / 2.08	0.00 / 0.16	1.00 / 0.66	0.47 / 2.53	– / 2.75	2.47 / 11.09
[30]	0.81 / 1.82	1.00 / 1.31	0.00 / 0.02	0.80 / 0.28	0.33 / 2.78	– / 0.52	2.94 / 6.73
[35]-1	-0.52 / 0.88	-0.33 / 1.65	-1.00 / 0.41	0.80 / 0.06	0.60 / 0.75	– / 1.05	-0.46 / 4.79
[35]-2	0.62 / 0.87	0.33 / 1.83	-1.00 / 0.53	-0.60 / 0.09	0.73 / 0.49	– / 1.46	0.09 / 5.27
[31]	-0.81 / 0.86	0.33 / 1.60	1.00 / 0.38	-0.20 / 0.63	-0.60 / 0.67	– / 3.33	-0.28 / 7.47
[42]	-0.81 / 1.37	0.67 / 1.91	1.00 / 0.31	0.40 / 0.35	0.47 / 1.00	– / 1.62	1.72 / 6.56
[41]	-0.43 / 1.57	-0.67 / 0.71	1.00 / 1.93	0.00 / 0.08	0.33 / 0.12	– / 0.24	0.24 / 4.65
[40]	-1.00 / 1.75	-1.00 / 2.15	-1.00 / 1.08	-0.40 / 0.34	-0.47 / 1.43	– / 5.55	-3.87 / 12.31
$\mathcal{C}_{CRE}$	0.90 / 1.72	1.00 / 2.10	1.00 / 0.99	0.60 / 0.26	0.47 / 0.86	– / 4.28	3.97 / 10.21
[28]	0.81 / 2.01	0.67 / 2.98	1.00 / 1.76	0.80 / 1.63	0.47 / 0.80	– / 6.04	3.74 / 15.23
[34]	-0.71 / 1.06	-0.33 / 0.84	0.33 / 0.24	0.40 / 0.21	0.33 / 0.80	– / 1.33	0.02 / 4.47
$\mathcal{C}_{\sigma}$	-0.81 / 1.44	0.00 / 1.29	0.33 / 0.17	0.40 / 0.17	0.33 / 0.61	– / 1.19	0.26 / 4.87
$\mathcal{C}_1$	0.05 / 1.31	-0.33 / 1.88	-0.33 / 0.10	-0.80 / 0.09	-0.73 / 0.45	– / 0.59	-2.15 / 4.41
$\mathcal{C}_{PC}$	0.62 / 1.19	0.67 / 1.49	1.00 / 0.20	1.00 / 0.14	0.60 / 1.13	– / 0.62	3.89 / 4.78
[36]	0.43 / 1.17	0.33 / 1.66	1.00 / 0.16	0.60 / 0.10	0.47 / 0.33	– / 0.92	2.83 / 4.33
[44]	-0.52 / 0.98	-0.33 / 1.37	-1.00 / 0.10	-0.60 / 0.25	-0.47 / 0.75	– / 1.42	-2.92 / 4.88
$\mathcal{C}_2$	-0.43 / 1.28	-0.33 / 1.41	-1.00 / 0.19	-0.80 / 0.18	-0.47 / 0.29	– / 1.02	-3.03 / 4.37
[43]-1	0.14 / 0.93	-0.33 / 1.59	-1.00 / 0.11	-0.80 / 0.20	-0.47 / 0.43	– / 1.00	-2.46 / 4.27
[43]-2	-0.05 / 1.12	-0.33 / 1.85	-1.00 / 0.13	-0.80 / 0.16	-0.60 / 0.43	– / 0.82	-2.78 / 4.51
MA	0.57 / 1.38	0.51 / 1.67	0.79 / 0.47	0.62 / 0.31	0.49 / 0.88	– / 1.88	2.11 / 6.59

collections. Considering that both [35]-1 and [35]-2 stem from the first two principal components of the same method, this discrepancy suggests that the segmentation task relates to shape complexity in a different manner than the GT considerations of the other collections.

2.4 Discussions

Three out of the four 3D methods, [32, 30] and [35]-1, perform consistently well, regardless of the base shape and noise type, and surpass any of the 2D methods for Collection 1, whereas, in the remaining two collections, 2D methods overtake. The performance of 2D measures are unexpected given that they do not make use of the full shape information. Though, we note that, the considered 3D methods did not make use of the full shape information either; some of them ([30, 35]) downsample the shapes in Collection 1 and all of them used voxelized shapes in processing Col-

Table 2.4: Kendall τ when the GT is the mean and the standard deviation of the number of segments of the human segmentations.

Method	$\tau_{\mu_{\text{cat}}}$	$\tau_{\mu_{\text{all}}}$	$\tau_{\sigma_{\text{cat}}}$	$\tau_{\sigma_{\text{all}}}$
[32]	0.148	0.346	0.072	0.234
[30]	0.041	0.354	0.018	0.203
[35]-1	0.110	0.105	0.055	-0.006
[35]-2	0.151	0.417	0.065	0.262
[31]	-0.022	0.251	0.013	0.138
[42]	0.061	0.401	0.027	0.202
[41]	0.140	0.375	0.087	0.167
[40]	-0.082	-0.458	-0.089	-0.282
$\mathcal{C}_{CRE}$	0.110	0.585	0.083	0.331
[28]	0.132	0.600	0.117	0.350
[34]	0.131	0.671	0.066	0.283
$\mathcal{C}_{\sigma}$	0.075	0.540	0.077	0.244
$\mathcal{C}_1$	-0.037	-0.255	-0.037	-0.078
$\mathcal{C}_{PC}$	0.099	0.464	0.078	0.217
[36]	0.041	0.326	0.029	0.123
[44]	-0.105	-0.486	-0.056	-0.219
$\mathcal{C}_2$	-0.112	-0.501	-0.084	-0.227
[43]-1	-0.069	-0.372	-0.085	-0.168
[43]-2	-0.079	-0.395	-0.090	-0.188
MA	0.092	0.416	0.065	0.206

lection 2 and 3. Nevertheless, it's worth emphasizing that the leading 2D methods vary across different tasks within Collections 2 and 3, with none excelling across all tasks.

An interesting observation is that [28], [40] and [34] perform poorly under the changes of the parameter w despite their high scores under the changes of the parameter c. We believe this could be because adjusting the width parameter w has a greater impact on the local changes in curvature than the parameter c, suggesting that these methods are inherently curvature-dependent. Similar results for the methods that measure convexity can be explained in a similar manner since convexity can be related to curvature for the examples in our datasets.

The results for Collection 1 suggest the use of different methods for different use

cases. For example, [32], [35]-1 can be used in applications involving additive perturbations and [31] can be used in applications involving noisy spheres. Provided one has information about the type of the noise present in their use cases, one can settle for [28], [40], $\mathcal{C}_2$, [43], or [44]. For overall robustness, [30] can be preferred.

The results for Collection 2 suggest that classical measures supported by psychology experiments are better alternatives for quantifying perceptual complexity as judged from the final product of the design process (*i.e.* ignoring the complexity in the generation process).

For Collection 3, we observe that the performances of the methods improve significantly when the entire collection is considered. In this sense, we can say that, for shapes from the same category, it is harder to correlate the shape complexity with the standard deviation or number of segments.
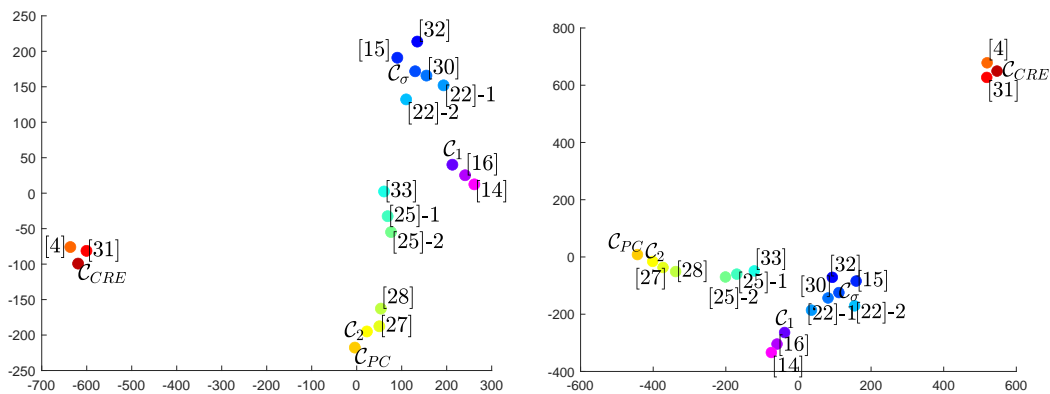
Figure 2.5: Clustering of the methods in τ-based feature space: 2D embedding applied to 17 (left) and 21 (right) τ-values

In Figure 2.5, two 2D embeddings of the evaluated complexity measures using Stochastic Neighborhood Embedding (t-SNE) [47] are depicted. For each measure, a feature vector is formed using the Kendall rank correlation coefficients reported in Tables 2.1-2.4. For the plot on the left, 17-dimensional feature vectors (whose components are the twelve τ scores from Table 2.1 and five τ scores from Table 2.3) are used. For the plot on the right 21-dimensional feature vectors are used by augmenting the 17-dimensional vectors with four additional τ scores from Table 2.4. We considered the scores from Collection 3 as optional as we consider the nature of this collection to be

22

different from the other two based on interpretation of the results. While acquiring the t-SNE plot, we negate the τ scores of the methods, [40], $\mathcal{C}_1$, [44], $\mathcal{C}_2$, [43]-1 and [43]-2 as they serve as measures of simplicity rather than complexity. The perplexity parameter is set to 2 for both plots. Nevertheless, we observed that doubling or even quadrupling the perplexity parameter does not make a significant qualitative change except that the spread gets larger.

Notice that [28], [40] and [34] form a distinct cluster. An interesting observation from Figure 2.5 is that the two methods [31] and [32], both employing real valued fields computed using a common partial differential equation, are not close in the τ-based feature space. This can be due to the difference of the methods in the metric spaces they each employ (L^2 vs. L^∞). The choice of the metric makes [32] an ideal method for noisy cubes whereas [31] is better suited for noisy spheres.

In addition to the correlations between GT and the measures, we find that the cor-

| Additively perturbed cubes | Additively perturbed spheres | Subtractively perturbed cubes | Subtractively perturbed spheres |

| Collection 2 Family 1 | Collection 2 Family 2 | Collection 3 Category-wise | Collection 3 Over all shapes |
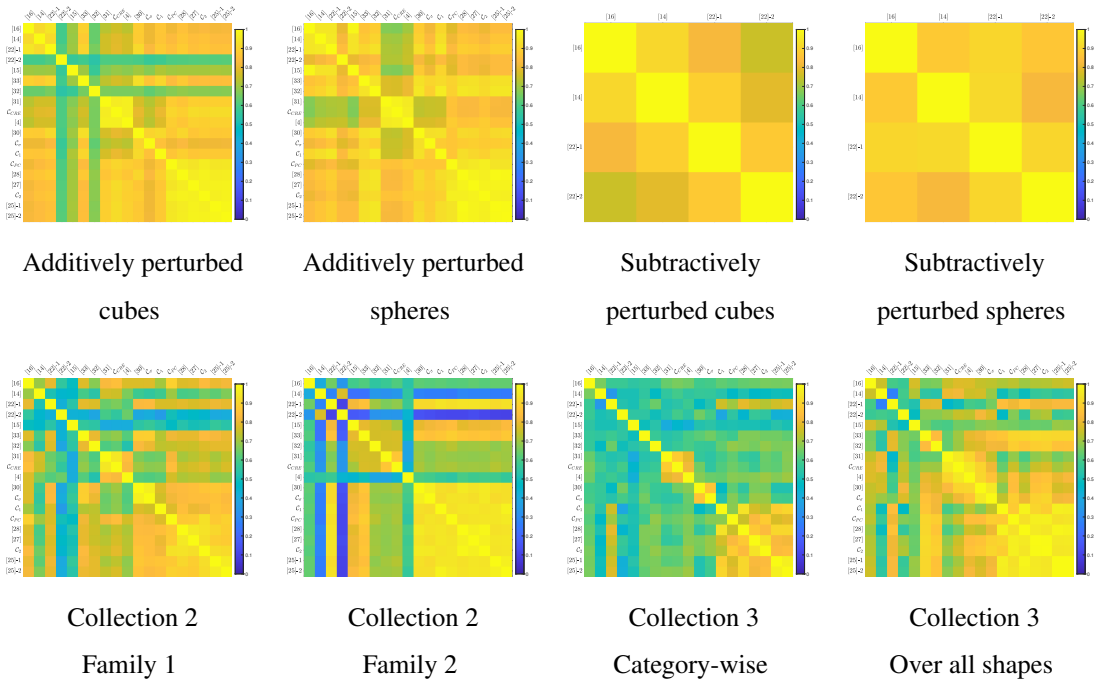
Figure 2.6: Correlations between pairs of methods

relations among the measures convey insight into the shape complexity. Thus, we compute and report in Figure 2.6 the Kendall rank correlation coefficients (τ) for each method pair. Specifically, for Collection 1 we compute the mean of τ_i ($i \in$

23

$\{1, 2, \ldots, 50\}$) for the groups of each family, for Collection 2, we compute τ over the families (*i.e.* disregarding the groups), and for Collection 3 we compute both the mean of τ_{cat} over the categories and τ_{all} over the whole collection. Here also we negate the scores of the methods, [40], C_1, [44], C_2, [43]-1 and [43]-2. The results show that the methods correlate the most with each other on the additively perturbed spheres. One possible explanation is that the diverse approaches to measuring shape complexity, such as uniformity of curvature, convexity, or the agreement of the shape with the underlying grid, more or less align for this family. Similar clusters to the ones seen to emerge in Figure 2.5 can be identified, such as the cluster composed of [28], [40] and C_{CRE}, or the cluster consisting of the convexity-based measures. Yet, for example, the second family of Collection 2 provides a means of distinguishing [28] from C_{CRE} and [40]. The same family also allows us to differentiate between the behaviors of the 3D methods. Similarly, the within-category results acquired on Collection 3 show that [44] and C_2 are closer to each other than they are to [43]-1 and [43]-2, and vice versa. Taken together, these provide further support for our claim that the three collections account for different aspects of shape complexity.

The correlations for Collection 3 are generally lower than those for Collections 1 and 2. This could be a consequence of either the data being more challenging, or that the proxy task does not relate strongly to complexity. This needs further study, and can be explored in future work.

CHAPTER 3

ENCODING SHAPE COMPLEXITY VIA SCALAR FIELDS

This chapter is based on our paper [37]. In order to maintain the integrity of the thesis, slight modifications are made to the content of the published paper. With similar concerns, some parts of the paper are moved to Chapter 7 and our further experimentations with the proposed method are presented in Appendix A.

Our aim in this work is to show how complexity can be encoded in a distributed manner, *i.e.*, as a scalar field, over a shape. We further attempt to convey the semantical meaning of the complexity-encoding field by means of quantitative and qualitative considerations.

3.1 Introduction

Given an 8-connected digital binary pattern representing a digital shape as a mapping $\mathbb{Z}^2 \to \{0, 1\}$, we are interested in quantifying, at each point on the pattern, the likelihood that the point belongs to a maximal *prototype* shape that fits the digital shape represented by the binary pattern in question. For the prototype shape themselves, the constructed scalar field is expected to be uniformly zero over the shape. The prototype shape serves as the simplest shape in a certain context.

The practical use of such a measure is two fold: First, if integrated over the pattern, the resulting number can be used as a measure of the tileability of the shape by the maximal prototype shape, which in turn can be used to quantify shape's complexity. Second, directly as a local measure, it can be useful in identifying the locations to cut the shape so that the resulting pieces represent an unrefined segmentation with

respect to the prototype. A perfectly tileable shape can be digitally represented with maximum compression. By *local culprits*, we mean the points where the value of the field deviates significantly from zero, as they are the points causing non-tileability.

In this work, we focus on the case where the reference shape is a rectangle. Nevertheless, by changing the underlying metric, the method naturally extends to the case where the prototype is a diamond (*cf.* Figure 3.9). The further extension of the method to much more general reference shapes is established in Chapter 4. We illustrate the method on 2D shapes which are not necessarily simply connected. Yet, all discussions are valid in higher dimensions.

3.1.1 Measures of rectangularity

Quantifying rectangularity has practical uses in several applications e.g. urban planning and landscape ecology [48]. Rectangularity measures are also used to improve over-segmented images [49]. In the literature, there are several global measures for quantifying the conformity of a shape to a simple prototype [50, 51, 32]. These global measures do not convey point-wise information. For circular shapes defined in $\mathbb{R}^2$, the method in [52] provide local information for quantifying conformity to circles.

A related problem of recent interest is quantifying the complexity of high-dimensional datasets for estimating their classification difficulty. Varshney and Willsky [6] measured their classifiers in terms of level sets of the decision hypersurfaces' geometrical complexity using ϵ-entropy. A growing number of works emphasize the role of the shape of the decision hypersurface as a determinant of either how complex the data is or how robust its classification by a certain classifier. An interesting claim by Fawzi *et al.* [53] is that vulnerability to adversarial attacks is related to positive curvature of the decision boundary. Fawzi *et al.* further attributed the robustness of the popular deep networks to the flatness of the shape of the produced decision boundaries.

Our numerical construction relies on the connection between differential operators and shape sets Chapter 7. Specifically, we resort to applying morphological derivatives to numerically approximate the infinity-Laplacian as in [54]. Proper numerical realizations of PDEs mimicking morphological process is an important issue. Among

the recent works is [55] where the flux-corrected transport scheme to the PDE imple-
mentation of erosions and dilations with arbitrary structuring elements is considered.

3.2   Method

For a shape S, we consider the following PDE with the infinity Laplacian:

$$\left(\Delta_\infty - \frac{1}{\rho^2}\right) f_S = -1 \text{ subject to } f_S\Big|_{\partial S} = 0. \tag{3.1}$$

While acquiring solutions to (3.1), ρ is chosen to be equal to the shape radius, *i.e.*
the maximal value of ℓ^∞ (chessboard) distance transform. After obtaining f_S, it is
normalized such that the maximum value of the field is 1. These ensure the scale in-
variance of f_S. To acquire numerical solutions, the approximation to Laplace operator
in L^∞ [54],

$$\Delta_\infty f_S(x) \approx \max_{y\in B(x)} f_S(y) + \min_{y\in B(x)} f_S(y) - 2f_S(x) \tag{3.2}$$

is used where $B(x)$ denotes a unit ball centered at x,. The numerical solution to (3.1)
can be acquired by using the scheme proposed in [32].

The equation (3.1) is favorable because the level curves of f_S roughly serve as gradual
transformation of the shape boundary ∂S towards a square under the influence of
the diffusion governed by the ℓ^∞ metric. The points at a system governed by (3.1)
generate and cumulate the values of the field, f_S. For squares, due to their isotropy
in ℓ^∞, the total accumulated values of points equidistant from the boundaries are the
same. Therefore, for a square S, the value of the field at x depends only on the
minimum distance of x to boundary (equivalently, on its distance to the *shape center*,
by which we mean the points attaining the maximum distance transform value). The
equidistant points form an equivalence class. As a result, the problem of acquiring
an analytic solution reduces to disjoint one dimensional problems over regions of the
square, which are continuous on the intersection of the regions.

In view of these, analytical solutions to Eq. (3.1) for up-right squares are acquired as

explained in Chapter 7. The analytical solution can be described as

$$f_S(x, y) = \rho^2 - \rho^2 \frac{e}{e^2 + 1} \times$$
$$\left(\exp \left\{ \frac{\max \{|x|, |y|\}}{\rho} \right\} + \exp \left\{ \frac{- \max \{|x|, |y|\}}{\rho} \right\} \right). \tag{3.3}$$

Although the solution is derived for a square, it naturally extends to rectangles. This is because the equivalence classes of a square, which are again squares, deform to rectangles: the distances to boundary and the shape center still add up to ρ since the shape center is a line for a rectangle rather than a single point. The validity of the acquired solution for the elongated unit-circle (rectangle in this case) is due to ℓ^∞ norm, and in the general scheme does not hold. For example, in the case of ℓ^2, an analytical solution for which is given in [52], circles are the corresponding equivalence classes, yet, the solution for circles does not apply for ellipses.

In the present form the solution is not translation invariant. To make it so, implicit reference to the origin should be removed. This can be done by reformulating (3.3) in terms of ℓ^∞ distance transform since $\max \{|x|, |y|\} = \|(x, y)\|_\infty$. We acquire:

$$f_S = \rho^2 - \rho^2 \frac{e}{e^2 + 1} \left(\exp\{t'^\infty\} + \exp\{-t'^\infty\} \right) \tag{3.4}$$

where $t'^\infty = 1 - t^\infty/\rho$, and t^∞ refers to ℓ^∞ distance transform of S.

In Figure 3.1 (b), the difference between the normalized (*i.e.*, has 1 as its maximum value) numerical solution $\hat{f}_{S,numerical}$ and the normalized analytical solution $\hat{f}_{S,analytical}$ for a square of side length 256 is displayed.

(a) $\hat{f}_{S,numerical}$ (b) $\hat{f}_{S,numerical} - \hat{f}_{S,analytical}$
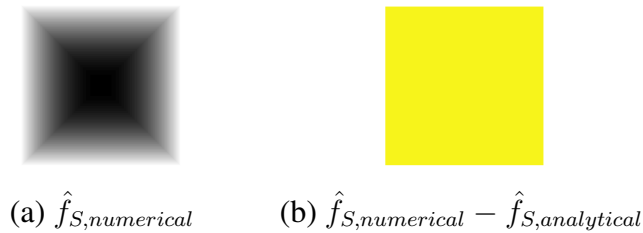
Figure 3.1: Field illustrations on a square of 256×256 pixels.

The maxima of the non-normalized fields are 5766.3 and 5797.8, respectively. The point-wise mean error between the normalized fields is

$$E = \frac{\int |\hat{f}_{S,numerical} - \hat{f}_{S,analytical}|}{|S|} = 0.001$$

This is an acceptable deviation, taking into account that numerical solution is acquired up to a first order approximation of Δ_∞.

To construct a measure valid for all shapes, we need (3.4) to extend beyond rectangles. Thankfully, the solution is in terms of the distance transform of the shape and can be deployed as is as an extension of the field. Then for any shape S,

$$f_{S,assumed} := \rho^2 - \rho^2 \frac{e}{e^2 + 1} \left(\exp\{t'^\infty\} + \exp\{-t'^\infty\} \right).$$

With this choice, we can assign scores of contribution to complexity to each point in the shape by simply subtracting the assumed extension from the numerically acquired solution. Thus we define the complexity encoding field d_S as

$$d_S := \hat{f}_{S,numerical} - \hat{f}_{S,assumed}.$$

This corresponds to measuring the error in assuming each point is coming from a square of radius ρ in which the point is located $\rho\, t'^\infty$ away from the center.

Acquired fields, $\hat{f}_{A,numerical}$ and d_S, for a square with an appendage of size 64×128 on one side are shown in Figure 3.2.

(a) $\hat{f}_{A,numerical}$ (b) d_S
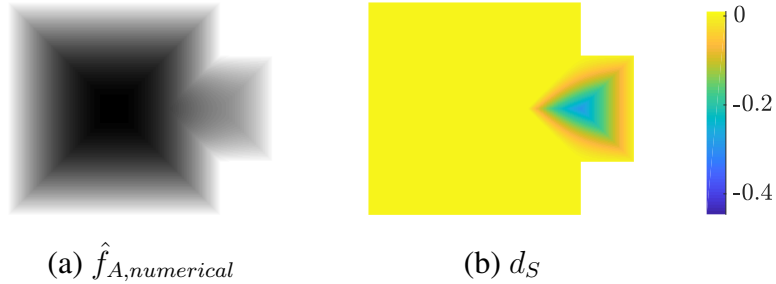
Figure 3.2: Field illustrations for a square with a rectangular appendage.

The introduced protrusion, being one of the simplest ways of increasing complexity, is informative in understanding the behavior of the proposed field. Looking at Figure 3.2 (b), we observe that the negative values occur around the rectangular appendage. Pixels near boundary, be them of base square or appendage, attain smaller values and would be disregarded in a thresholded treatment of the field. The extrema is attained exactly at the two center pixels in the vertical direction along the edge of the square.

3.3 Illustrative Experiments

In all the experimental results depicting d_Ss, if a color bar is not shown, we use a color scale is between 0 and -0.446 where the zero is denoted by yellow. In the first experiment, we demonstrate the method on composite rectangles of constant width. The size of the square is 64×64, *i.e.*, one quarter of the previous square used in Figure 3.1. As shown in Figure 3.3, these shapes have no pixels that increase the complexity. Again, slight fluctuations in d_S are due to the first order approximation to the infinity Laplacian operator.
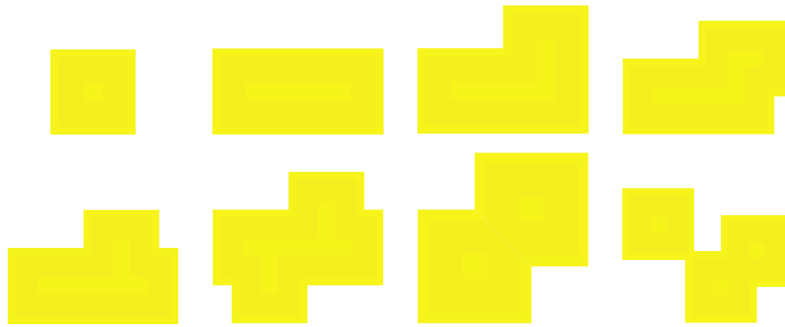
Figure 3.3: Composite rectangles of constant width

In the second experiment, we apply the method to floor plans of increasing complexity. Results are depicted in Figure 3.4. The first floor plan is composed of four identical rooms. This plan has no pixels that increase the complexity. As we introduce missing or extra segments, respective locations start to attain negative d_S values. The last floor plan consists of multiple rooms of varying sizes. The two rooms of the largest size are deemed as the simplest with d_S values near zero. d_S attain higher negative values inside the smaller rooms.

Note that the scalar field d_S is parameterless. As such, it implicitly measures the deviation from the rectangle that maximally fits into the shape due to the choice of ρ (§ 3.2). Hence, smaller rooms are identified as parts of the plan that increase the plan's complexity.

For specific purposes, however, one might be tolerant of the size or interested in identifying complexity with reference to a prototype shape of certain size rather than the maximal size. To this end, ρ can be treated as a parameter of d_S. In that case,
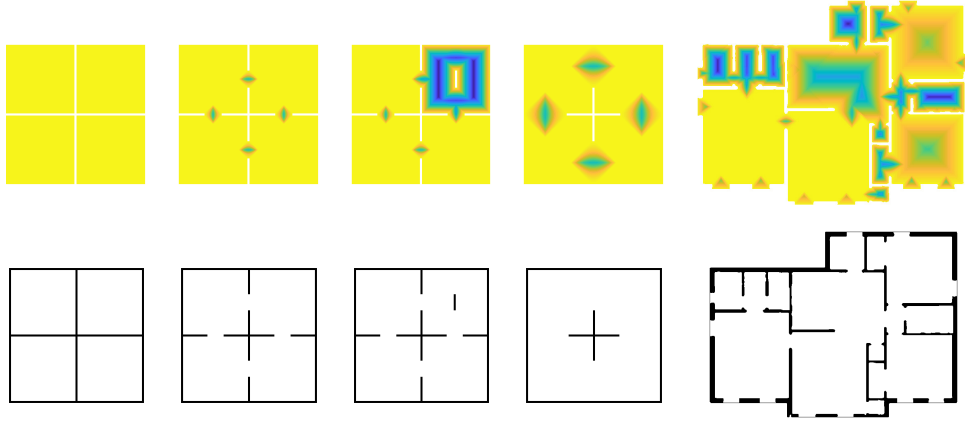
Figure 3.4: d_Ss (top row) for some floor plans (bottom row)

positive values for d_S arise in the central parts of larger rectangles. This is illustrated in Figure 3.5. The rightmost figure shows the original result from Figure 3.4, *i.e.*, $\rho' = \rho$. In the remaining two results, notice that d_Ss take both negative and positive values as indicated by the color bars. In the leftmost figure, $\rho' = 8\rho/9$ coinciding with the width of the two identical square-shaped rooms on the right-side of the plan. Inside these two rooms d_S is almost uniformly zero as indicated by the color bar. Furthermore, the number of pixels where d_S is negative decreases.

$\rho' = 8\rho/9$ $\qquad\qquad$ $\rho' = 4\rho/9$ $\qquad\qquad$ $\rho' = \rho$
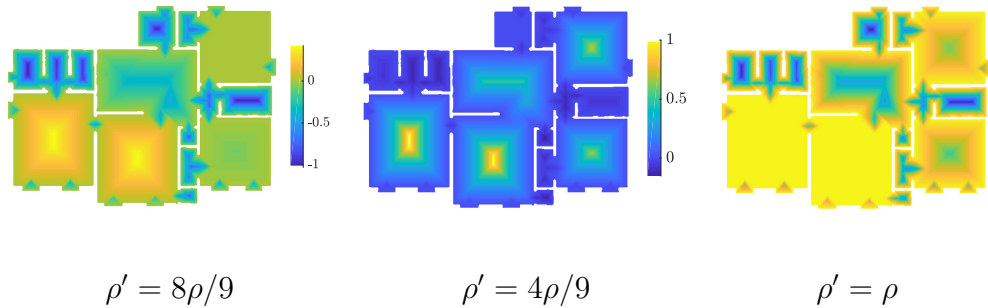
Figure 3.5: When the critical width ρ' is treated as a parameter

In the third experiment, we explore quantifying the shape complexity with a single value by using $\text{mean}(|d_S|)$ as the complexity indicator. Some illustrative results are shown in Figure 3.6.

In the final experiment we explore how we can simplify a complex shape towards a rectangle. We had observed that the local extrema of d_S are near the centers of regions that increase complexity. Furthermore, at the extrema, the gradient of $f_{S,numerical}$
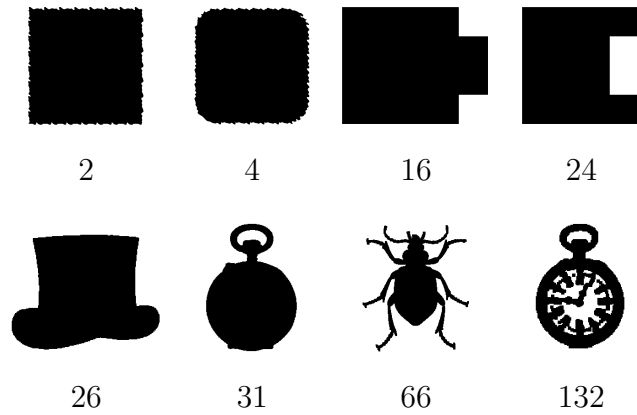
Figure 3.6: Sample complexity quantifications for shapes.

indicates the position of such regions relative to the shape body. Thus, the orthogonal direction to the gradient reveals the directions for cuts that lead to a more rectangular shape.

As a proof-of-concept, we used a greedy iterative algorithm. It leverages both fields at each iteration step, d_S providing information on the cut location and $f_{S,numerical}$ providing information on the cut direction. For example, for the square with an appendage (Figure 3.2), the mean gradient of $f_{S,numerical}$ at the two neighboring extrema of d_S has no component in y direction. Therefore, the shape is cut along the y direction from these extrema, separating the base square and the appendage.
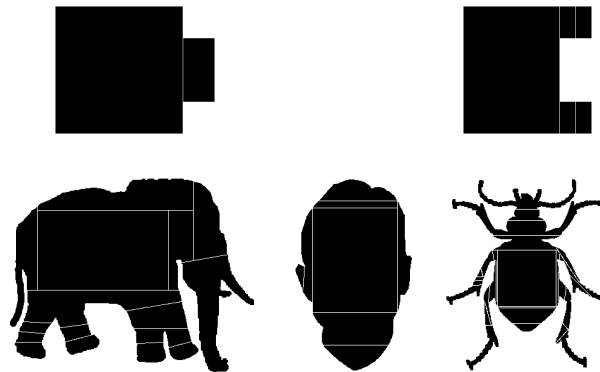
Figure 3.7: Iterative simplifications of various shapes towards a rectangle (best viewed in digital).

Illustrative cuts are shown in Figure 3.7. When determining the direction normal to the gradient, the numerically computed direction is replaced with the direction of

closest axis if the angle between them is lesser than 2.86 degrees ($\approx \arctan 0.05$).

In Fig. 3.8 the mean absolute value and maximum absolute value of the proposed fields for squares with increasing appendage sizes are shown. The appendage sizes are $k \times 2k$ pixels for $k \in \{0, 4, \ldots 128\}$ where the base square is of 256×256 pixels. Note that this means the initial shape is a square and the final shape is a rectangle. Hence, the two ends of the scores attain the same mean and maximum scores as expected. The extremum of the maximum absolute value plot is attained for the appendage of size 60×120. The mean absolute value plot is skewed due to the increasing number of non-zeros present in the constructed fields, which in turn is due to the increasing area of the appendages.
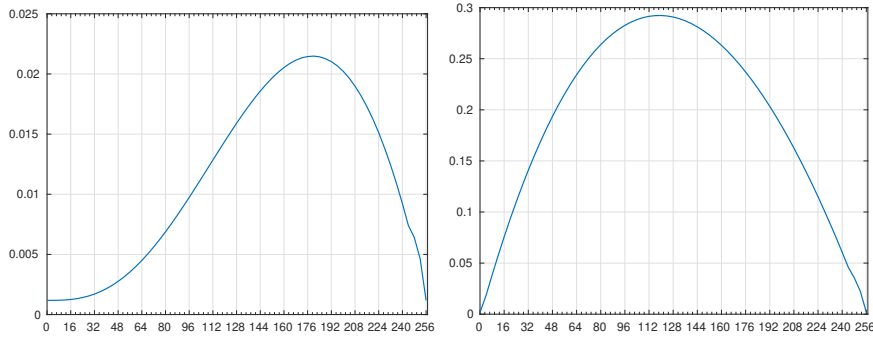
Figure 3.8: Mean absolute values and maximum absolute values of the complexity-encoding fields for squares with increasing appendage sizes

Throughout the chapter, we assumed that the prototype is a rectangle (or hyper-cuboid in higher dimensions). Yet, similar results can be obtained by using the unit balls of ℓ^1 and the corresponding distance transform; the analytical solution Eq. (3.4) extends seamlessly to that case. See Figure 3.9 for sample results where we demonstrate measuring rhombicity.

We proposed a new measure d_S for distributed coding of the shape complexity. Each pixel on the shape is assigned a value d_S quantifying the pixel's contribution to overall complexity where the simplicity is understood as the expressibility of the shape in terms of an assumed prototype. We performed proof-of-concept experiments to demonstrate the practical applicability of the method. In the next chapter, we extend our method to be compatible with more general prototypes.
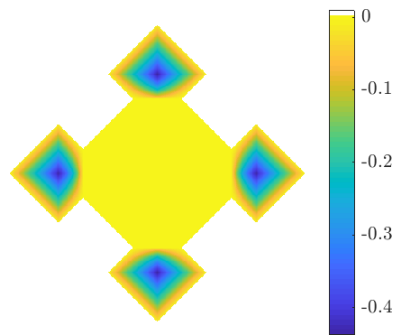
Figure 3.9: Extending the method to measure rhombicity

CHAPTER 4

RELATIVE SHAPE COMPLEXITY

In this chapter, to the question of "what is the complexity of a given shape?", *i.e.*, the problem of quantification of shape complexity, we answer with a question:

What is considered as the simplest shape?

Our point in this answer is that complexity is intrinsically relative, and it is challenging to discuss the term without a reference point.

Consequently, we propose a measure of relative shape complexity. Proof-of-concept results are presented in this chapter. Applications for deep learning are presented in Chapter 5, where we achieve state-of-the-art-results, and Chapter 6, where statistically significant improvements on several computer vision tasks are acquired, to further establish the use of the proposed method. But first, a discussion about the relative aspect of complexity is in order.

4.1 The Relative Aspect of Complexity

The lack of a canonical definition of complexity is mentioned often in the literature. We attribute a part of why this is so to the relative aspect of complexity. Gell-Mann, in [2], states that measures that relate to complexity are relative as they are dependent on *i)* the previous knowledge, *ii)* the idealization of the deployed computational model's capabilities, *iii)* the language of description, and *iv)* the level of detail used in describing the object. The following argumentation is based on several different examples of complexity, by means of which, we elaborate particularly on the first two factors. Thereby, we attempt to confine what we mean by the word "relative" in our

35

context.

In linguistics, the complexity of a language, as in the level of intricacy or difficulty of that language, can be quantified in different ways [56]. The measures are broadly referred to as measures of linguistic complexity. The complexity of a sentence can be perceived differently depending on the observer's linguistic background. For example, in [57], researchers show that the first language (L1) background of participants contributes to the complexity of the second language (L2), and in [58], the morphological complexity of L1 of the participants is found to play a crucial role in estimating their difficulty while acquiring Dutch as L2. Similarly, [59] reports that the syntactic complexity of English writings of non-native speakers (NNS) differentiate from those of native speakers when the L1 background of NNS are taken into account. Thus, the previous knowledge is an important factor while quantifying complexity.

Limiting ourselves to the theory of computation for mathematical rigor, we consider the *complexity of computable problem*s [60, 61]. Each computable problem, denoted as P, admits a class of computable solutions that can be represented as algorithms ($\{A_i^P\}_{i \in \mathbb{N}}$). It is common convention in the theory of computation to consider the scaling of the needed resources of an algorithm under changes of problem's input size as the algorithm's complexity. For example, when the resource under consideration is time, we are talking about computational complexity ($CoC(\mathcal{A}_i)$), when it is space, space-complexity ($SpC(\mathcal{A}_i)$). In a similar vein, we can talk about the complexity of a problem as the complexity of the optimal algorithm that solves it, *i.e.*, $CoC(P) = \min_i\{CoC(\mathcal{A}_i^P)\}$, $SpC(P) = \min_i\{SpC(\mathcal{A}_i^P)\}$.

With these established, we consider a specific problem. Our interest in this problem is that it demonstrates a change of complexity classes with a change in the assumptions about the nature of the allowed operations on data.

Example 4.1.1. (the Deutsch-Jozsa problem [62]) Let $f : \{0,1\}^n \to \{0,1\}$ be a function such that is either constant or balanced (for exactly half of the inputs it evaluates to 0). Determine whether f is constant or balanced.

Any classical solution to this problem has to evaluate the function $2^{n-1}+1$ times in the

worst case scenario, *i.e.*, in exponential time. However, the Deutsch-Jozsa algorithm solves this problem in polynomial time (in fact, using a single evaluation of f [63]), demonstrating that the mathematically rigorous notion of computational complexity is relative to the allowed operations on data.

Another example conveys the message more clearly: with a compass, it is easier to draw a circle than a triangle; with a ruler, vice versa. Thus, more generally, we say complexity is relative to the available tools.

Returning back to shape complexity, we argue that it is also relative. Our notion of relativeness is with respect to the prior knowledge represented via predetermined simple shapes. As mentioned before, most measures of shape complexity *implicitly* takes circles as the simplest shapes. The measure $P^2/4\pi A$ is a canonical example. Our aim is to make this implicit assumption *explicit*. In the next section, we start by defining a mathematically rigorous and sensible notion of a family of simplest shapes determined by what we refer to as *prototype* shapes.

4.2 Defining Simplest Shapes

A common theme in complexity studies is the notion of compressibility. If the object at hand is highly compressible, it is considered simple, otherwise, more complex.

Example 4.2.1. (Complexity and compressibility)

1. Kolmogorov-Chaitin complexity [64, 65]: One can measure the complexity of a given object as the length of the shortest algorithm, written in a predetermined programming language, that precisely reproduces it. This complexity measure incurs that the more compressible an object is, the lower its complexity. For example, a compressible string has low Kolmogorov-Chaitin complexity due to the availability of a short description for it.

2. Linguistic complexity: The structural complexity of a language [66] can be measured by the number of grammatical structures/rules in it. In a sense, the

37

number of different structures supported by a language indicates its compressibility.

We follow a similar logic by letting a shape in an embedding space be a *prototype*[1] and defining the shapes acquired by a certain operation to be the corresponding family of the simplest shapes [2]. The operation we consider is the successive morphological dilations of a single point with this prototype. In this way, the prototype acts as a compressed representation of this family with respect to morphological dilation (*cf.* § 4.5).

Ideally, our definition of simplest shapes is generic enough to hold for shapes of arbitrary dimensionality in both continuous and discrete domains.

Example 4.2.2. (Simplest shapes in $\mathbb{R}^n$)

In the following, we indicate the names of the corresponding boundaries in parentheses.

1. (Hypercubes) Let $\mathbb{C}_\epsilon = \{x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n : \|x\|_\infty \leq \epsilon\}$ denote a closed n-cube of side length $2\epsilon > 0$ in $\mathbb{R}^n$. Then, all n-cubes $\mathbb{C}_{k\epsilon}$, where $k \in \mathbb{Z}^+$, are included in the family of simplest shapes with respect to $\mathbb{C}_\epsilon$ in view of

$$\mathbb{C}_{(k+1)\epsilon} = \mathbb{C}_{k\epsilon} \oplus \mathbb{C}_\epsilon.$$

2. (Hyperspheres) In the same grounding as the n-cubes, the closed n-ball of radius $\epsilon > 0$, $\mathbb{B}_\epsilon = \{x \in \mathbb{R}^n : \|x\|_2 \leq \epsilon\}$, admits all n-balls $\mathbb{B}_{k\epsilon}$ where $k \in \mathbb{Z}^+$ in its family of simplest shapes.

3. (Hyperellipsoids) More generally, the closed n-ellipsoid with $a = (a_1, a_2, \cdots, a_n)$, $\mathbb{E}_{\epsilon, a} = \{x \in \mathbb{R}^n : \| (x_i/a_i)_{i=1}^n \|_2 \leq \epsilon\}$, admits all n-ellipsoids $\mathbb{E}_{k\epsilon, a}$ in its family of simplest shapes.

In the discrete setting, such as the infinite rectangular grid, similar, yet, different families of simplest shapes can be defined. The main point of difference is that some shape

---

[1] These are the structuring elements in mathematical morphology, *cf.* § 4.5. We prefer to use the word "prototype" instead, to emphasize the semantic context we assign to these objects.
[2] We consider shapes to form an equivalence class under spatial translations.

38

families that can be encoded with an arbitrarily small prototype can no longer be encoded in a discrete domain due to the *finite representation capability*. Most notably, objects that require arbitrary precision, such as hyperellipsoids, do not constitute a family of simplest shapes in the discrete domain[3]

Example 4.2.3. (Simplest shapes in $\mathbb{Z}^n$)

1. (Hypercubes) Let $\mathbb{C}_m = \{x = (x_1, x_2, \cdots, x_n) \in \mathbb{Z}^n : \|x\|_\infty \leq m\}$ denote an n-cube in $\mathbb{Z}^n$ of side length $2m + 1 \geq 0$ where m is an integer. Then, all n-cubes $\mathbb{C}_{km}$, where $k \in \mathbb{Z}^+$, are included in the family of simplest shapes with respect to $\mathbb{C}_m$.

2. (Hyper-diamonds) Similarly, $\mathbb{D}_m = \{x = (x_1, x_2, \cdots, x_n) \in \mathbb{Z}^n : \|x\|_1 \leq m\}$ admits all n-diamonds $\mathbb{D}_{km}$ in its family of simplest shapes.

3. Denote by $\mathbb{L}_m$ the n-cubic lattice in $\mathbb{Z}^n$ of length m in each direction and index the lattice sites by n numbers, each in the closed interval $[1, m]$, in increasing order. Then we define a regular n-checkerboard of diagonal length $(2m + 1)$ (denoted $\mathbb{H}_{2m+1}$) as a binary pattern on $\mathbb{L}_{2m+1}$ for which all the lattice sites with all of its coordinates being an odd number are 1, and the rest of the sites are 0. The family of simplest shapes of $\mathbb{H}_{2m+1}$ contains all n-checkerboards $\mathbb{H}_{2km+1}$.
[4]

Sample discrete shapes in two and three-dimensions are demonstrated in Figure 4.1.

In practice, prototype shapes of desired size can be attained from a given sample of a shape by means of rescaling. Figure 4.2 exemplifies this scenario. We acquire a mean prostate shape from the NCI-ISBI13 dataset, and obtain a prototype by rescaling it by 1/12 in each direction via nearest-neighbor interpolation. This approach can be considered as a lossy compression scheme. In the case depicted in Figure 4.2, a compression rate of $1.7 \times 10^3 \approx 12^3$ is achieved. The corresponding lossy reconstruction

---

[3] This is not due to our choice of definition for simplest shapes but is due to the discrete space structure. As shown by Rosenfeld [67], the same pathological situation occurs for the measure $P^2/4\pi A$. See [68] for a more elaborate discussion.

[4] The checkerboard in three-dimensions relates closely to the *face-centered cubic crystalline structure* in crystallography.
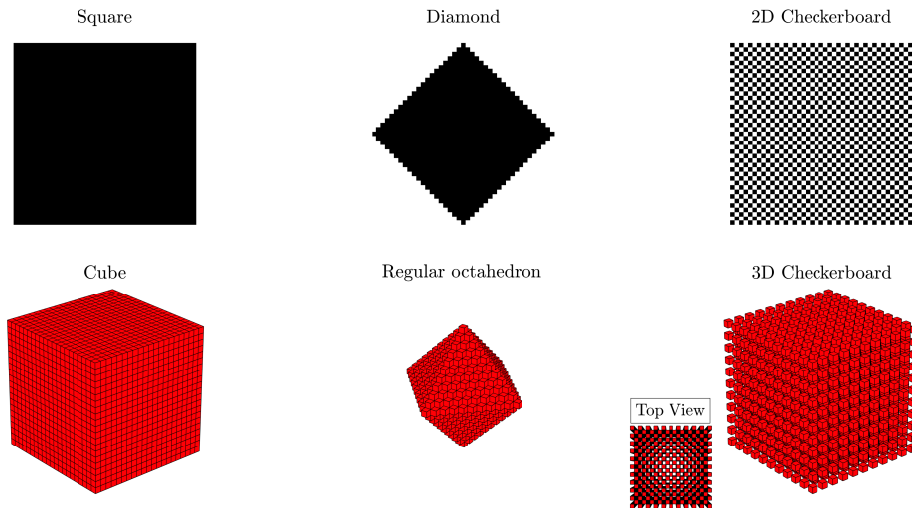
Figure 4.1: Simplest shapes in $\mathbb{Z}^2$ and $\mathbb{Z}^3$: (left) n-cubes, (middle) n-diamonds, (right) n-checkerboards.

is depicted in Figure 4.2 (c). Note that the lossy reconstruction can be considered as a polygonal approximation to the original shape.

(a) Mean prostate shape
($\approx$365.000 voxels)

(b) Prototype (1/12)
(213 voxels)

(c) Lossy reconstruction
($\approx$350.000 voxels)
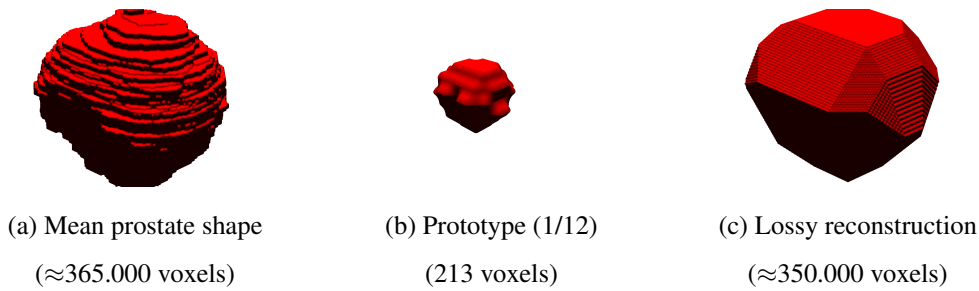
Figure 4.2: Prototypes can be acquired from a given shape by means of rescaling. This defines a lossy compression approach for shapes.

4.3 Measuring Relative Shape Complexity

So far, we have established the notion of the family of simplest shapes relative to a prototype. Our aim now is to construct a measure of shape complexity that is *i)* compatible with the defined notion of simplest shapes, *i.e.*, relative, *ii)* extends consistently under incremental deformations to other shapes, and *iii)* exactly computable in discrete domains.

Before introducing our measure, we first define the notion of a generalized distance transform (DT) that will be helpful in the following explanations.

Definition 4.3.1. (Generalized distance transform) The **generalized distance transform** (DT) $t_{S,s}$ of a shape S with respect to the prototype s encodes the distance of each point in S from the boundary and is acquired as the sum of successive erosions (*cf.* § 4.5) of S by s.

For prototypes such as a square or a diamond, the underlying metric spaces ℓ_∞ and ℓ_1 along with the corresponding DTs are well-known. However, an analytic expression of the DT for an arbitrary prototype is in general not available. Therefore, we construct the generalized DT $t_{S,s}$ as given in Alg. 1. Note that the generalized DT is an integer-valued function. We refer to the set $\{x \in S : t_{S,s}(x) = \max t_{S,s}\}$ as the *shape center*.

Algorithm 1 Generalized DT of a shape w.r.t. a given prototype

---
Ensure: S and s are binary images

$\quad t_{S,s} \leftarrow S$

$\quad$ **while** S is not empty **do**

$\qquad S \leftarrow S \ominus s$

$\qquad t_{S,s} \leftarrow t_{S,s} + S$

$\quad$ **end while**

4.3.1 Governing equation

We consider

$$\left(\Delta_s - \frac{1}{\rho^2} \right) f_{S,s} = -1, \text{ subject to } f_{S,s}\Big|_{\partial S} = 0 \tag{4.1}$$

where s is the prototype and Δ_s denotes the corresponding morphological Laplacian operator (*cf.* § 4.5). We choose the parameter ρ to be the maximum of the generalized DT of S with respect to s.

Note that, to a first order approximation, the infinity Laplacian Δ_∞ deployed in Chap-

ter 3 and the morphological Laplacian Δ_s are equivalent since [69]

$$\Delta_\infty v(x) \approx \max_{y \in B(x)} v(y) + \min_{y \in B(x)} v(y) - 2v(x)$$

on a regular grid where s is a square of unit diameter and $B(x)$ is the unit ball centered around x.

Migration from the infinity Laplacian to the morphological Laplacian reflects a change of perspective on digital shape complexity. Specifically, for exact computability, the deployed mathematical framework should have exact digital representations since the underlying space is digital. With this in mind, the morphological Laplacian is a natural choice, as it has a discrete representation that is compatible with the digital space. In contrast, the infinity Laplacian is a continuous operator by definition. Thus, in a discrete setting one can only approximate it. Therefore, by switching to the morphological Laplacian, we use tools that are well-suited to the space in which the objects of interest are embodied.

In Chapter 7, we show that analytical solutions to Eq. (4.1) can be obtained as a function of the generalized DT for the simplest shapes with respect to a prototype s as:

$$
\begin{aligned}
f_r &= A\lambda_1^r + B\lambda_2^r + \rho^2, \\
\lambda_{1,2} &= \frac{1 \pm \sqrt{1 - 4\gamma^2}}{2\gamma}, \\
A &= \frac{\rho^2 \left(\lambda_2^\rho - \frac{\rho^2}{\rho^2+1}\lambda_2^{\rho-1} \right)}{\lambda_1^\rho - \lambda_2^\rho - \frac{\rho^2}{\rho^2+1}\left(\lambda_1^{\rho-1} - \lambda_2^{\rho-1} \right)}, \\
B &= -A - \rho^2
\end{aligned}
\tag{4.2}
$$

where $\gamma = \frac{\rho^2}{2\rho^2+1}$ and f_r denotes the value attained by the elements of the set $\{x \in S : \min \text{dist}(x, \partial S) = r\}$, that is, the points that are r away from the nearest boundary.

4.3.2 Complexity-encoding field and measuring complexity

The solution given by Eq. (4.2) is exact only for the simplest shapes of the prototype. Nevertheless, interpreted purely as a function of the generalized DT, it can be generalized to arbitrary shapes, albeit no longer exact. Interpreted in this way, the function hypothesizes a form for the solution. We denote this function as $f_{S,s}^{\text{h}} = f_{S,s}^{\text{h}}(t_{S,s})$, "h"

42

standing for hypothesized. If S is a shape in the family of simplest shapes of s, then $f_{S,s}^{\mathrm{h}}$ coincides with the actual solution $f_{S,s}$. Otherwise, it only approximately satisfies Eq. (4.1). The disagreement between the hypothesized solution and $f_{S,s}$ increases as S becomes less regular with respect to the prototype. Therefore, we propose

$$\mathcal{C}_{S,s} = 1 - \frac{f_{S,s}}{f_{S,s}^{\mathrm{h}}}$$

as a **relative shape complexity-encoding field**.

Semantically, this field measures the error of $f_{S,s}(x)$ at each point $x \in S$ compared to the field value when x comes from a simplest shape of radius ρ in which x is located $\rho - t_{S,s}(x)$ away from the shape center.

We encode complexity not through a statistical tool, such as entropy, but rather through a hypothesis about what is simple. Hence, it is possible to use any measure of deviation between the hypothesized and constructed field as a single number indicating the complexity of a field. We opt for relative error:

$$c = \frac{\|f_{S,s}^{\mathrm{h}} - f_{S,s}\|_2}{\|f_{S,s}^{\mathrm{h}}\|_2}.$$

We also use $d = \mathrm{round}\left(c \times 10^3\right)$ when precision is not of primary concern.

In practice, $f_{S,s}$ is constructed using one of the numerical solution strategies outlined in Chapter 7.

4.4 Proof-of-Concept Results

In this section, we present our results that serve as a proof-of-concept. Initially, we consider results pertaining to the feasibility of the proposed method by showing that the theoretical simplest shapes are in fact quantified as such. Further, we show that the method passes this feasibility test for most non-convex prototypes. Next, we separately consider two directions for quantifying the complexity of arbitrary shapes: *i)* complexity with respect to a probe, *ii)* complexity in the sense of self-similarity. The former is useful when we are interested in quantifying the complexity of a collection of shapes given a building block. The latter relates the complexity of a shape to itself

and provides an absolute measure of complexity. In both instances, our experiments consistently yield meaningful results.

4.4.1 Feasibility of the proposed method

To establish the feasibility of the proposed method, we focus on the validity of the hypothesized solution for simplest shapes under changes of shape and size of the prototype. The only restriction we impose on prototypes is that each point belongs to its own neighborhood, *i.e.* the center element of prototype is nonzero.

Convex prototypes. Let $S_{s,k}$, $k \in \mathbb{Z}^+$, denote a simplest shape with respect to the prototype s such that $S_{s,1} = s$ and $S_{s,k+1} = S_{s,k} \oplus s$. For prototypes fitting into $2 \times 2, 2 \times 3, 3 \times 3$, and 3×4 grids, we tested the agreement between the hypothesized solution and the numerically acquired solution for $S_{s,k}$, $k \in \{1, 2, 10, 50\}$ for *all* possible prototypes. For all convex prototypes[5], the maximum complexity score is approximately 8×10^{-9} when the numerical field is initialized as zero. It's important to note that this non-zero value results from the convergence threshold of the numerical scheme, set to 10^{-8} in these experiments. In fact, if we initialize the field while acquiring the numerical approximation $f_{S,s}^{\mathrm{n}}$ with $f_{S,s}^{\mathrm{h}}$, the iterations converge at the 0-th step with $c = 0$. We report c for $f_{S,s}^{\mathrm{n}}$ initialized from zero to account for imprecisions arising from the numerical scheme.

Non-convex prototypes. For prototypes represented in $2 \times 2, 2 \times 3$, and 3×3 grids, $f_{S,s}^{\mathrm{h}}$ and $f_{S,s}^{\mathrm{n}}$ agree for all possible non-convex shapes as well. This includes prototypes such as the checkerboard pattern ▨[6]. We observed that for some non-convex shapes in 4×3 grid, the two fields agreed less; e.g. for prototypes with holes such as ▦ ($c = 4.7 \times 10^{-6}$ for $S_{s,50}$) and most notably for disconnected prototypes such as ▨ ($c = 1.29$ for $S_{s,50}$). For prototypes in larger grids, we performed non-

---

    [5] Although trivial, there is one exception: prototypes which encode a point as its only neighbor. In this case, $S_{s,k} = s$ for all k. Corresponding solutions can be acquired by changing the expression for f_ρ to take itself as its both minimum and maximum neighbor.
    [6] We represent prototypes in the grid form where white squares corresponds to the points included in the neighborhood of the center element of grid. The center element of an $n \times m$ grid is the element with coordinates $(\lfloor (n+1)/2 \rfloor, \lfloor (m+1)/2 \rfloor)$ in a coordinate system where the top left element has coordinates $(1, 1)$.

exhaustive tests and observed similar results.

(a) $c = 7.54 \times 10^{-9}$ (b) $c = 7.49 \times 10^{-9}$
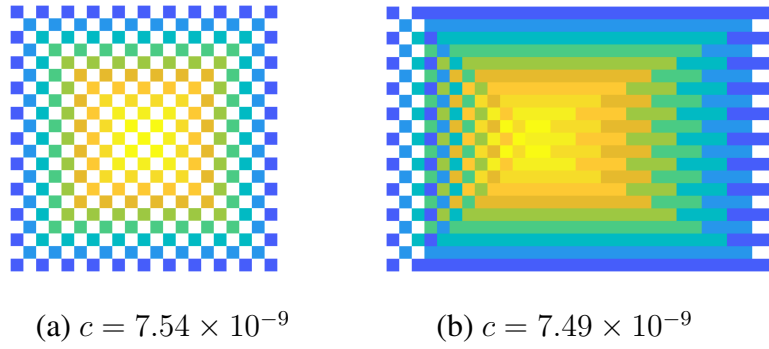
Figure 4.3: The solutions to Eq. (4.1) constructed on two unusual simplest shapes – $f^{\mathrm{n}}_{S,s}$ for $S_{s,10}$ of (a) $s = $ and (b) $s = $

Our findings suggest that $f^{\mathrm{h}}_{S,s}$ applies well beyond the prototypes chosen as the unit balls of the usual metrics, ℓ_∞ and ℓ_1. Feasible results might be expected even for non-convex prototypes (as in Figure 4.3), though disconnected prototypes should be approached cautiously.

4.4.2 Complexity with respect to a probe

Two-dimensional probes. Figure 4.4 displays the set of probes we use. We measure the complexity of a collection of shapes from MPEG7 dataset [70]. The complexity-encoding fields and complexity scores is shown in Figure 4.5 where within each quarter of the figure the used probe corresponds to the one in Figure 4.4. Each field is shown at its own scale, hence the used color map is to give only a sense of the high complexity parts within a shape. When comparing different shapes the reported d scores should be kept in mind.

(a)   (b)   (c)   (d)
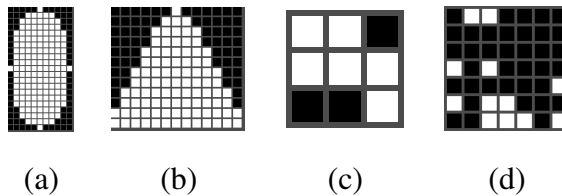
Figure 4.4: 2D prototypes used as probes.

The first probe is a digitized ellipse. The complexity measurements with respect to it yields the face and jar shapes as the simplest two, Figure 4.5 (a). The second probe mimics the rounded triangle shape which, indeed, attains the minimum score, Figure 4.5 (b). We include the third probe especially for its behavior on the star shape: it identifies the regions of the star with similar inclination to its simplest shape, Figure 4.5 (c). The fourth probe is acquired randomly as a sparse pattern in a 7×7 grid to demonstrate the behavior of the constructed field for different shapes for a random disconnected prototype, Figure 4.5 (d). Note that despite being disconnected and not including points in their own neighborhoods, the probe yields consistent results with respect to a scheme where the simplest shape is the convex hull of the probe.

For the first three probes, in each shape, an area roughly corresponding to the maximally fitting up-right ellipse/triangle/inclined region is considered to be the least contributor to the complexity. For the fourth probe we observe a similar behavior not with respect to the probe itself, but to its convex hull. We emphasize that this is despite not only the non-convex but also the disconnected nature of the fourth probe.
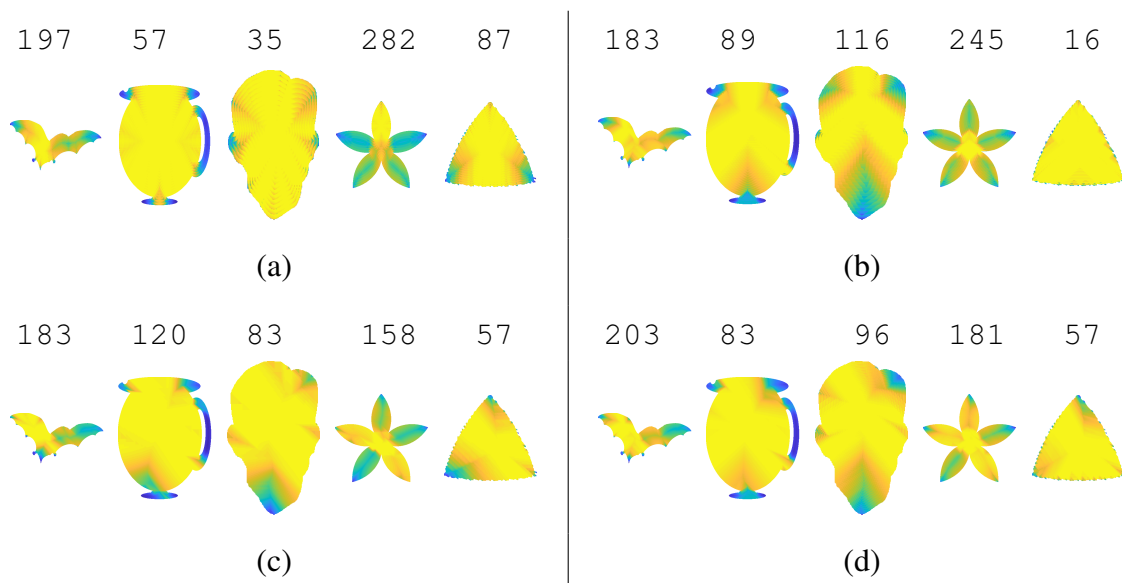
Figure 4.5: Five shapes probed with prototypes from Figure 4.4. The numbers on top of each shape is its complexity score (d value).

Three-dimensional probes. For our experiments in three dimensions, we first consider a selection of shapes from the QSC dataset [11]. The used probe and its simplest

shape ($S_{s,10}$) are shown in Figure 4.6.   The collection of shapes along with their d

Probe $c = 7.29 \times 10^{-9}$
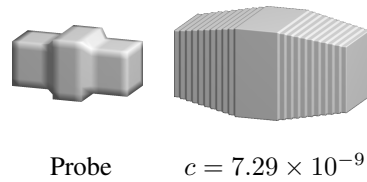
Figure 4.6: A 3D probe s and its corresponding isotropic shape $S_{s,10}$.

scores with respect to the probe are displayed in Figure 4.7. We observe that the more convex the shape gets, the more similar it becomes to the simplest shape of the probe, hence, attaining lower complexity scores.

$d = 24$ $d = 29$ $d = 32$ $d = 35$ $d = 44$
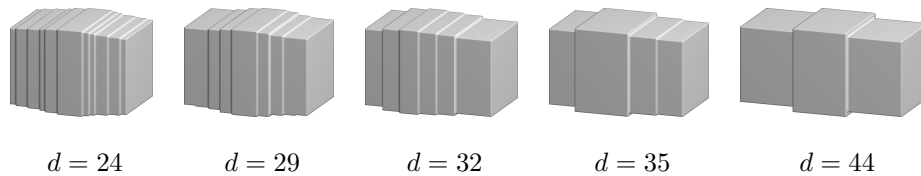
Figure 4.7: Sample results on shapes from the QSC dataset.

Secondly, we consider shapes from Princeton segmentation benchmark [46]. Each shape displayed in Figure 4.8 is voxelized to have approximately 1.6×10^6 voxels. We probe each shape in Figure 4.8 with a sphere prototype fitting into a $7 \times 7 \times 7$ grid. Corresponding complexity scores are reported in the figure as d_2.

$d_1 = 46$ $d_1 = 73$ $d_1 = 127$ $d_1 = 160$ $d_1 = 246$ $d_1 = 300$

$d_2 = 97$ $d_2 = 73$ $d_2 = 129$ $d_2 = 133$ $d_2 = 172$ $d_2 = 224$
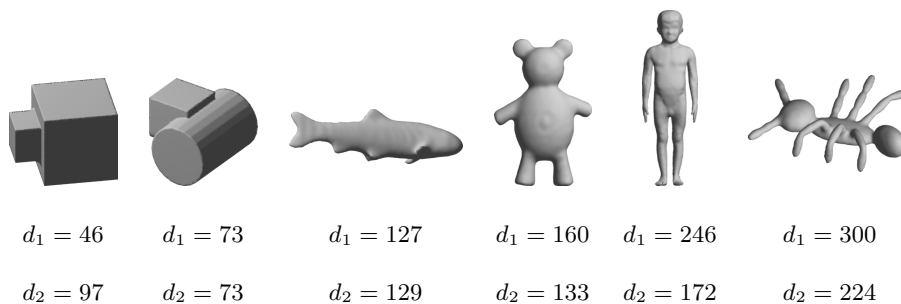
Figure 4.8: Sample results in 3D. d_1 is the complexity score in the sense of self-similarity at the scale $1/20$, d_2 is the complexity score with respect to a spherical prototype.

4.4.3 Self-similarity

To measure complexity in the sense of self-similarity, we acquire a prototype from each shape by rescaling the shape. The prototype is used as a probe for the shape itself. At a given scale this serves as an absolute measure of complexity.

In Figure 4.9 we present results for two collections from the MPEG7 dataset. The prototype for each shape is acquired by rescaling the shape by $1/35$. Note that for

(a)   (b)
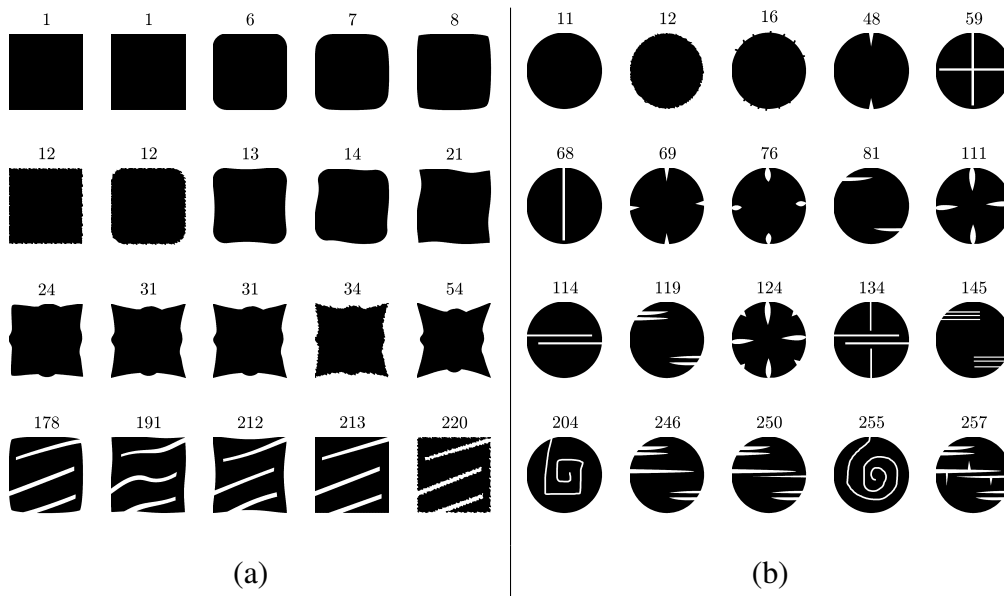
Figure 4.9: The complexity-ordered collections (a) device3 and (b) device9 from MPEG7 with respect to prototypes acquired by re-scaling the original images.

both basic shapes, fairly low complexity scores ($d = 1$ for a square and $d = 11$ for a circle) are acquired. Further, we observe that structural differences such as those in boundary curvature or number of spikes are accounted for by our measure. Finally, we notice that the highest increases in the complexity of shapes in these collections are due to indentations, and that closer they are to the center of the shape, the more impactful they get.

The reported d_1 scores in Figure 4.8 reflect the complexities when the scale parameter is $1/20$. Here, we observe that the simplest shape in the sense of self-similarity is the cube with an appendage and the most complex is the ant shape owing to its high non-convexity.

48

4.5 Some Definitions in Mathematical Morphology

Definition 4.5.1. (Morphological dilation and erosion) The **morphological dilation**, $v \oplus s$, of a set (or function) v with respect to the **structuring element** s is the operation $\oplus$ such that

$$(v \oplus s)(x) = \sup_{y \in s} v(x - y).$$

The dual transformation, **morphological erosion**, $v \ominus s$ is given by

$$(v \ominus s)(x) = \inf_{y \in s} v(x + y).$$

Morphological dilation and erosion are dual operators in the sense that, for binary sets v,

$$v \oplus s = \left(v^c \ominus s \right)^c, \qquad v \ominus s = \left(v^c \oplus s \right)^c,$$

where the superscript c denotes the set complement operation.

Definition 4.5.2. (First-order morphological differential operators) The **internal gradient** operator, denoted ∇_s^-, is defined by its action as

$$\nabla_s^- f(x) = f(x) - (f \ominus s)(x).$$

Similarly, the **external gradient** operator, denoted ∇_s^+, is defined as

$$\nabla_s^+ f(x) = (f \oplus s)(x) - f(x).$$

Definition 4.5.3. (Morphological Laplacian) The **morphological Laplacian**, denoted Δ_s, is a second-order morphological differential operator, that can be expressed conveniently as a combination of the external and internal gradients,

$$\Delta_s f(x) = (\nabla_s^+ - \nabla_s^-) f(x) = (f \oplus s)(x) + (f \ominus s)(x) - 2f(x).$$

The action of the morphological differential operators are exemplified in Figure 4.10. Notice the change in the edge thickness and detected edges (for example, around the sideburns) when the structuring element is altered.

Development of mathematical morphology started with the works of Matheron and Serra. A chronological and insightful first-hand account of the birth of mathematical morphology is given in [71]. Matheron and Serra's works on binary sets was
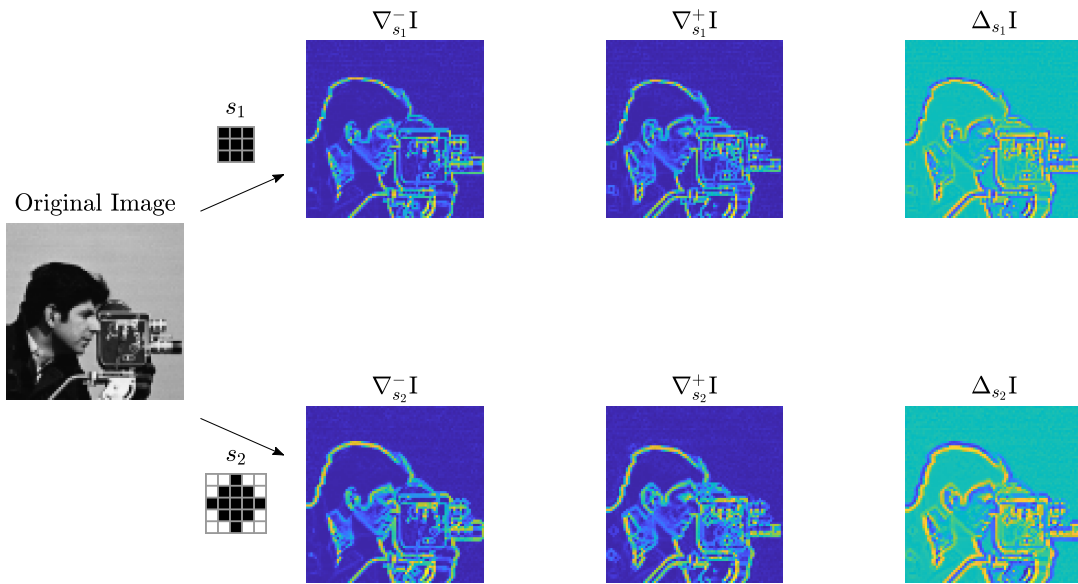
Figure 4.10: The effect of morphological differential operators (internal gradient ∇_s^-, external gradient ∇_s^+ and the morphological Laplacian $\Delta_s = \nabla^+ - \nabla^-$) with respect to two different structuring elements (s_1 and s_2) on the image "Cameraman".

followed by the development of gray-scale morphology. Nakagawa and Rosenfeld [72] are the first to consider such a generalization. This is followed by the works of Sternberg [73] and Serra [74]. Heijmans [75] studied the gray-scale morphological operators in detail. Morphological differential operators such as external/internal gradients (also called sup/inf-derivatives in 1D [76]) and morphological gradient [77], and morphological Laplacian [78] have been introduced and found applications in image segmentation [79], image coding [80], processing tensor-valued data [81].

CHAPTER 5

APPLICATION: GENERALIZABLE MEDICAL IMAGE SEGMENTATION VIA THE RELATIVE COMPLEXITY ENCODING FIELD

In this chapter, we propose a novel application for shape complexity. We build upon our earlier claim that the proposed shape complexity-encoding field is semantically rich, and demonstrate one way it can be employed in deep learning. The specific problem we consider is single-source domain generalization for prostate segmentation.

In the domain generalization (DG) scenario, a model is trained on a set of data from *source domain*(s), and tested on previously unseen data from *target domain*s. The case when the model is trained on a single source domain is called as the single-source DG scenario.

In medical imaging, training data is scarce (due to privacy policies and costly annotation processes) and of highly heterogeneous characteristics (due to different protocols, magnetic field strengths, etc.). Thus, neural networks with large number of parameters, whose performance is highly dependent on the diversity of training data, are trained by multiple passes over a small set of data. Consequently, despite achieving decent performance on the source domain, models trained in this manner tend to overfit to the training set, and, when tested on target domains, the performance decreases drastically.

Our initial results are very promising, achieving state-of-the-art performance in prostate segmentation.

5.1 Related Work

We briefly review domain generalization, domain adaptation, and explicit usage of shapes in deep learning.

5.1.1 Domain generalization and adaptation

DG methods can be broadly classified into three main categories: data-based, learning-based, and representation-based approaches.

Data-based approaches aim to ensure generalizability by means of careful augmentation of the data to span a larger distribution during training. In BigAug [82], the authors take a pipelined approach and show that transformations relating to image quality, appearance, and spatial configuration can be applied in a stochastic and stacked manner to obtain models that are robust under domain shifts. This simple yet effective approach constitutes a strong baseline for DG methods. More recently, [83] considered nonlinear filtering and transformations for augmentation with the aim of disentangling image content from image style. The works [84, 85] target augmentations using intensity shifts. While [84] achieves this via global intensity nonlinear augmentations (GIN), [85] formulates the intensity transformation as a mapping that can be learned using adversarial training to produce samples that challenge the network. In a similar vein, adversarial domain synthesizer (ADS) [86] employs two adversarial branches that integrate the approach of the GIN and introduces a separate branch to ensure the synthesized images do not alter the semantics of the source image.

Descriptive statistics of available data can also be considered as a means for DG. For instance, in the training phase of [87], adaptive instance normalization (AdaIN, [88]) is employed to counteract the bias towards global statistics. The authors adopt a stochastic application of AdaIN at the batch level in this approach. In the context of federated learning, [89] incorporates AdaIN, allowing clients to share descriptive statistics of their individual data, effectively tackling privacy concerns. An alternative perspective involves treating feature statistics as non-deterministic values derived

from an underlying probability density function (PDF). This is exemplified in [90], where the authors assume that feature statistics follow a Gaussian distribution.

The spectral properties of available data is also utilized for data augmentation. In [91] and [92] explore generalizable learning based on the phase spectrum of data. Their approach is based on the observation that the amplitude and phase spectra of the the Fourier transform reflect features with distinct semantic levels. Specifically, the amplitude spectrum is associated with low-level features like style, while the phase spectrum carries higher-level semantics, such as the shape of an object. To exploit this property, they employ amplitude spectrum mixing which aims at training models that are robust under changes of features with low-level semantics.

Learning-based approaches, on the other hand, explore different learning strategies that warrant generalizable feature learning. Meta-learning is used by [93, 94] assuming the existence of data from multiple source domains. In their framework, the realistic domain shift scenario is mimicked during training by treating some of the source domains as meta-test and the rest as meta-train sets. EpiFCR [95], with more training costs, proposes to leverage dedicated feature-extracting and task-specific networks for each domain that facilitate in training the main network. The auxiliary networks are utilized only during training and discarded afterwards. This is similar to [96] where additional classifiers are used for entropy regularization.

In addition to these approaches, the gradient vectors during training are also utilized to acquire generalizable models via favoring the learning of features with lower gradients [88], or inter-domain gradient matching [97].

Representation-based approaches explicitly focus on the extraction of domain-invariant features. In DICA [98], the authors propose a kernel-based optimization algorithm maximizing the across-domain similarity while retaining the data-label relations. This approach shares similarities with dimensionality reduction algorithms such as kernel PCA [99], transfer component analysis [100], and covariance operator inverse regression [101]. For DA and DG, scatter component analysis (SCA) is proposed in [102] which combines the ideas from DICA, KPCA and kernel FDA [103].

SCA quantifies separability of data and classes, and across-domain dissimilarity in terms of a geometrical measure that they name as scatter in reproducing kernel Hilbert space. Representing the feature data with X and the labels with Y, the assumption that the marginal distribution $P(X)$ changes while the conditional probability distribution $P(Y|X)$ remains intact, shared by the methods mentioned so far, is suboptimal. Based on causality, a more realistic assumption can be made in which both $P(Y)$ and $P(X|Y)$ changes, where Y is assumed to be the cause of X. For example, [104] extracts conditional transferable components for which $P(h(X)|Y)$ is invariant where h is a transformation of data. Similarly, with both the marginal and conditional probability distributions changing, [105] proposes learning domain-invariant class conditional distributions. This is achieved by maximizing the between-class scatter while minimizing the scatter of class prior-normalized marginal distributions, total scatter of class-conditional distributions, and within-class scatter. In a similar spirit, [93], aims at retaining global inter-class relationships by aligning confusion matrices while locally, sample clustering is leveraged to ensure the compactness of extracted features. Some overarching ideas in [105] are implemented via adversarial learning in the deep learning setting in [106].

If data from multiple domains are available, requiring that domains cannot be identified looking at the latent representations alone effectively implies domain-invariant feature extraction. Two works that take this approach are [96] and [107]. Both works employ a domain discriminator. In the case of the former the model is trained in an adversarial manner, whereas in the latter work, domain discriminator is trained with randomized or wrongly-assigned labels.

UndoBias [108] explicitly addresses dataset bias by introducing a model that learns both a domain-invariant weight vector and a set of domain-specific bias vectors. MTAE [109] uses auto-encoders to account for inter-domain variability of object appearances, facilitating the learning of robust features. In a similar setting, [110] introduces a low-rank constraint on latent representations to ensure domain invariance of extracted features. Building on the observation that input data from different domains should share the same representation if derived from the same object, [111] proposes an object-conditional (as opposed to class-conditional) objective via matching-based algorithms.

Domain adaptation (DA), similar to DG, aims at increasing the performance of models on target domains. It differs from DG in that, during or following the training on source domain, the target domain data is leveraged to reduce the effect of distribution shift. For medical imaging, a comprehensive review of DA methods is [112]. Roughly, we can divide DA methods into two categories as latent space [113, 114, 115, 116, 117, 118, 119, 120, 121] and image space DA methods [122, 123, 124]. The latent space methods aim at aligning latent representations of source and target domains, whereas the image space methods aim at image translation from source to target domains. Making use of task-specific decision boundaries, [114] tries to align source and target domains. In [115], cross-domain feature matching via adversarial training is proposed. Addressing a shortcoming of adversarial learning, [119] takes category-level joint distributions into account for feature alignment. Hybrid methods, such as [125, 126, 127], utilize from both feature and image space alignment approaches. For example, in [127], the proposed method for DA is composed of a generation stage and an adaptation stage. In the generation stage image space data is utilized while in the adaptation stage feature-level adaptation is used.

5.1.2 Shapes in deep learning

Shape of an object is a domain-invariant feature. Hence, shape learning is appealing for both DG and DA, and utilized by many works [94, 128, 129, 130, 131]. In the context of DG, shape dictionaries and a shape coefficient predictor are utilized in [128] for segmentation. The study [94] proposes a shape-aware approach that aims at ensuring compact shape predictions while also enforcing accurate predictions near boundaries. BayeSeg [129] uses shape information by modeling shapes in terms of their spatial dependencies. In [130], the learning of structural properties are highlighted by means of structured edges. In the DA setting, [131] explores the usage of evolving shape priors (ESP). In their approach, using self-supervised learning, gradually refined labels for target domain data are generated. ESP is initialized as the mean shape of the source labels.

While CNNs excel at learning textures, they may not optimally utilize shape information [132, 133]. Therefore, leveraging auxiliary mechanisms to help CNNs in learn-

ing shape information is not limited to DG and DA, but of broader interest. Explicit methods, such as those using active shape or contour models, shape dictionaries, or Markov/conditional random fields, are reviewed in [134]. Methods that implicitly integrate shape information into end-to-end network training include those using shape transforms, variational autoencoders, and topological constraints. The most relevant to us among these are the methods that use shape transforms.

In the segmentation literature [135, 136, 137, 138, 139, 140, 141, 142], the most commonly used shape transforms are distance transform (DT) and signed distance transform (SDT). DT and SDT, despite encoding redundant information (in the sense that they can be acquired trivially from shape boundary alone), provides a means for forcing the networks to learn spatial relations within shapes. For the unbalanced segmentation problem, [135] defines a loss term based on SDT given by $\mathcal{L} = \frac{1}{|\Omega|} \sum_{\Omega} SDT(y) \otimes \hat{y}$ where Ω is the volume, $y, \hat{y}$ are the GT label and network prediction, respectively, and $\otimes$ denotes the Hadamard product. In [136], minimization of Hausdorff distance is considered where the loss term $\mathcal{L} = \frac{1}{|\Omega|} \sum_{\Omega} (y - \hat{y})^2 \otimes (DT^2(y) + DT^2(\hat{y}))$ is employed. On top of segmenting the target organ, DT/SDT prediction can be leveraged as a related task to ensure learning of spatial shape features more explicitly [138]. An additional prediction head is introduced to the segmentation network in [140, 141] to predict DT. Introducing more parameters, [142] proposes using a reconstruction branch for DT that shares only the encoding layers.

5.2   Proposed Approach

By their design, CNNs are better at learning texture rather than shape, as highlighted in [132, 143, 133]. Therefore, the success of a CNN crucially relies on guaranteeing that the correct texture is learned. However, two main issues in prevent this:

1. inaccurately annotated segmentations may force networks to learn incorrect textures,

2. the tissue of a patient may have atypical textures.

In both cases, the generalizability of the model is compromised when attempting to

56

learn incorrect or exceptional textures.

To address these issues, we propose a novel approach: weighting the importance of different parts of the GT label based on their contribution to the overall shape complexity. Our underlying assumption is twofold. First, we assume that the irregular parts of a patient's organ are more likely to have atypical textures. Second, we assume that the mislabelled boundaries increase shape complexity over that region.

To this end, after constructing complexity-encoding fields, we define importance masks as a linear function of the former. We use these importance masks to define an importance weighted Dice loss, Dice_{IW}, described in detail in the following.

5.2.1   Importance masks from complexity-encoding fields

Given that the complexity-encoding field $\mathcal{C}(S)$ of a shape S attain larger values on complexity increasing parts whereas simple parts attain a value near 0, we seek a function that weighs these regions otherwise. We opt for a simple choice and formulate f as a linear function of $\mathcal{C}$:

Definition 5.2.1. Let $\mathcal{C}_{\max}^{(d)}$ be defined as $\mathcal{C}_{\max}^{(d)} \coloneqq \max_{S \in \mathcal{D}}\{\max_{x \in S} \mathcal{C}(S)(x)\}$, *i.e.*, the maximum value of the complexity-encoding fields across all shapes in domain $\mathcal{D}$. Then, we call

$$f^{\alpha}(\mathcal{C}) \coloneqq (1 - \alpha) \times \frac{\mathcal{C}_{max}^{(d)} - \mathcal{C}}{\max_{x \in S}\{\mathcal{C}_{max}^{(d)} - \mathcal{C}\}} + \alpha \mathbb{1}, \qquad \alpha \in [0, 1]$$

as the **importance mask** of shape S in domain $\mathcal{D}$ relative to the prototype acquired from $\mathcal{D}$, where $\mathbb{1}$ is a constant field of 1s of appropriate dimensionality and size.

The parameter α controls the impact of complexity field on the importance mask; for $\alpha = 0$, $\mathcal{C}$ contributes maximally, and for $\alpha = 1$, $\mathcal{C}$ does not contribute at all. Although it is possible to update this parameter during training based on training progress or current loss, we prefer simplicity and use it as a constant. In our experiments, $\alpha = 0$ led to performance decrease on the test of the source domain. As a result, we used $\alpha = 0.1$.

5.2.2   Dice loss with importance weighting

Dice loss is a popular loss function for segmentation tasks, given as

$$\text{Dice}(\hat{y}, y) = 1 - \frac{2 \sum_{x \in \Omega} \hat{y}(x)\, y(x)}{\sum_{x \in \Omega} \hat{y}(x) + \sum_{x \in \Omega} y(x)}.$$

Formulated in this way, the loss function does not prioritize any pixels; all pixels are considered equally important.

To prioritize the learning of pixels that do not increase the complexity of the shape, we propose a modified Dice loss. Our modified Dice loss can be written as

$$\text{Dice}_{\text{IW}}(\hat{y}, y, C(S); f^\alpha) = 1 - \frac{2 \sum_{x \in \Omega} \hat{y}(x)\, y(x)\, f^\alpha(C)(x) + \varepsilon}{\sum_{x \in \Omega} \hat{y}(x)\, f^\alpha(C)(x) + \sum_{x \in \Omega} y(x)\, f^\alpha(C)(x) + \varepsilon}$$

or, more compactly,

$$\text{Dice}_{\text{IW}}(\hat{y}, y, C(S); f^\alpha) = 1 - \frac{2\, \hat{y}_i\, y_i\, f_i^\alpha(C) + \varepsilon}{\hat{y}_i\, f_i^\alpha(C) + y_i\, f_i^\alpha(C) + \varepsilon}$$

where we use Einstein's summation convention, *i.e.*, repeated indices are summed over. The variable ε is used following the common practice of "smoothing" the Dice loss. We use $\varepsilon = 10^{-5}$.

The modified Dice loss weighs the contribution of each voxel (pixel in the 2D case) of the shape based on the value of $f^\alpha(C)$ there. Effectively, this means that parts that are deemed to be complexity-increasing contribute less to the loss function. The gradient of Dice_{IW} with respect to the predicted variables

$$\frac{\partial}{\partial \hat{y}_j} \text{Dice}_{\text{IW}} = \frac{-2\, y_j\, f_j^\alpha(C)}{\hat{y}_i\, f_i^\alpha(C) + y_i\, f_i^\alpha(C) + \varepsilon}$$
$$+ f_j^\alpha(C)\, \frac{2\, \hat{y}_i\, y_i\, f_i^\alpha(C) + \varepsilon}{(\hat{y}_i\, f_i^\alpha(C) + y_i\, f_i^\alpha(C) + \varepsilon)^2}, \qquad j \text{ fixed}$$

showcases the effect of $f^\alpha(C)$ during optimization explicitly.

Example 5.2.2. In the one-dimensional setting, let $y = \{y_i\}_{i=1}^4 = (0, 1, 1, 0)$ be the GT, $\hat{y} = \{\hat{y}_i\}_{i=1}^4 = (0, 0.6, 1, 0.4)$ be a prediction, and $f(C) = \{f_i(C)\}_{i=1}^4 = (0, 1, 0.5, 1)$ be the corresponding importance weights. In this example, $f(C)$ indicates that the first parameter can be ignored and the third parameter is less important than the remaining two. In this case, whereas the conventional Dice loss yields

$$\text{Dice}(\hat{y},\, y) = 0.2, \qquad \frac{\partial}{\partial \hat{y}_i} \text{Dice}(\hat{y},\, y) = (0.2, -0.3, -0.3, 0.2),$$

the modified Dice loss gives

$$\text{Dice}_{\text{IW}}(\hat{y},\, y) = 0.27, \qquad \frac{\partial}{\partial \hat{y}_i}\text{Dice}_{\text{IW}}(\hat{y},\, y) = (0,\, -0.42,\, -0.21,\, 0.24).$$

The loss increased since the prediction on the more important y_2 is wrong. Similarly, gradient vectors have changed according to the relative importance of each parameter.

The modified Dice loss is very similar to the that employed in [144]:

$$1 - \frac{1}{C}\sum_c \frac{2\sum_i w_{ci} y_{ci} \hat{y}_{ci} + \epsilon}{\sum_i w_{ci}\left(y_{ci} + \hat{y}_{ci}\right) + \epsilon}.$$

In their case, w_{ci} are determined by the network's performance during training as $w_{ci} = \alpha|\hat{y}_{ci} - y_{ci}| + (1 - \alpha)$ where the subscript ci denotes corresponding quantities relating to the class c for voxel i. Their loss aims at emphasizing the learning of hard-to-guess voxels. Our method differs in that the importance weights are predetermined and introduce complementary information to the network. This additional information aids in reducing the learning of ungeneralizable features.

We mark that the design of the importance mask implies unequal treatment of patient data during prostate segmentation predictions. Prostates with more complex shapes (with respect to the prototype) receive less attention overall.

As described in §5.3, the network we employ makes background and prostate segmentation predictions simultaneously. Of these, the importance mask is utilized only in the calculation of the loss of prostate predictions. That is, we treat all background information as equally important, regardless of the prostate shape.

5.3 Experimental Setup

In our experiments, we use T2-weighted prostate MRI data from different sources following previous works [94, 86, 129]. The sources, RUNMC and BMC, are from NCI-ISBI13 dataset [145], HCRUDB is from I2CVB dataset [146], and UCL, BIDMC and HK are from PROMISE12 dataset [147]. The data specification of each is given in Table 5.1.

In Figure 5.1, the spectral properties of each domain is shown, demonstrating the inhomogeneity across the considered datasets.

Table 5.1: Specifications about the employed prostate MRI data from public datasets.

Dataset	Institution	Number of patients	Field strength	Resolution (in-/through plane) (mm)	Endorectal coil	Manufacturer
Site A	RUNMC	30	3T	0.6-0.625/3.6-4	Surface	Siemens
Site B	BMC	30	1.5T	0.4 /3	Endorectal	Philips
Site C	HCRUDB	19	3T	0.67-0.79 /1.25	No	Siemens
Site D	UCL	13	1.5T or 3T	0.325-0.625 /3-3.6	No	Siemens
Site E	BIDMC	12	3T	0.25 /2.2-3	Endorectal	GE
Site F	HK	12	1.5T	0.625 /3.6	Endorectal	Siemens
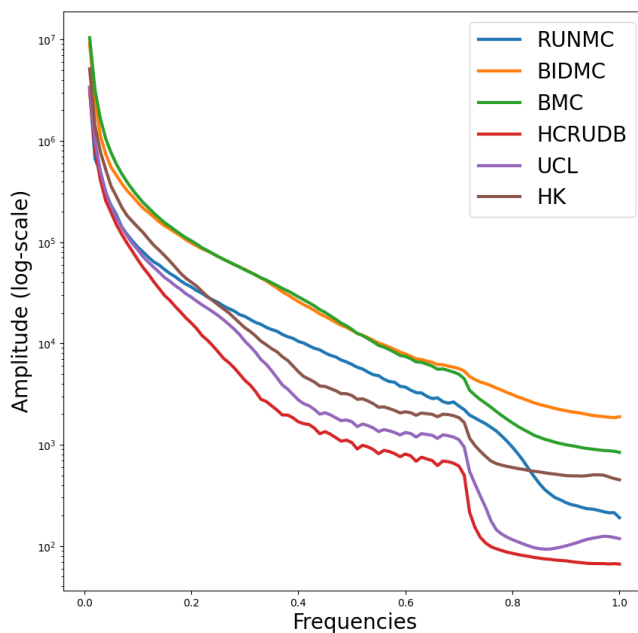
Figure 5.1: Amplitude (log-scale) versus frequency for prostate MRIs from three datasets and a total of six domains

5.3.1 Constructing the complexity fields

As a relative measure of shape complexity, our method has to be supplied with a reference shape, *i.e.*, a prototype. We acquire prototypes from the mean shapes by means of rescaling (as in Chapter 4). For completeness, we re-present the related visuals in Figure 5.2. In acquiring mean shapes, we first resample each patient's GT label to have isotropic spacing in all dimensions (each voxel represents a volume of $0.5208^3 mm^3$) and center-align each. Then, we select for the mean shape the voxels

that contribute to at least 60 percent of the samples.

(a) Mean prostate shape
(≈365.000 voxels)

(b) Prototype (1/12)
(213 voxels)

(c) Lossy reconstruction
(≈350.000 voxels)
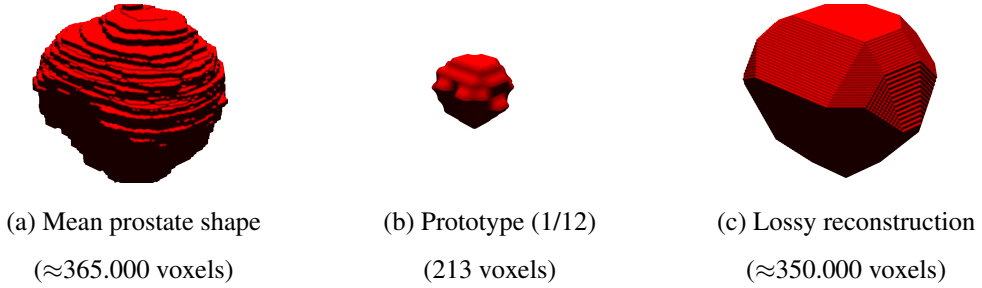
Figure 5.2: (a) Mean prostate shape acquired from training samples of RUNMC dataset (see § 5.3) (b) Prototype shape acquired by rescaling via nearest-neighbor interpolation (c) Lossy reconstruction of the mean prostate shape.

After the prototypes are acquired, we construct the 3D complexity-encoding fields as described in Chapter 4. It is worth to repeat that during the scalar field construction we only make use of the prototype and GT labels, and not the MRI data. In Figure 5.3, some slices from the constructed complexity encoding fields are shown. We observe that the constructed field attains large values mostly near the boundaries. Additionally, our observations indicate that regions of GT with atypical prostate texture appear to coincide with parts that contribute to an increase in complexity, supporting our initial assumption.

5.3.2   Training details

In the following, when applicable, the used method for interpolations of the MRI data and complexity-encoding fields is bilinear interpolation, and for the GT label, nearest-neighbor interpolation.

Preprocessing.   We perform data preprocessing through the following pipeline: each volumetric data is oriented to posterior left superior (PLS) orientation, rescaled to have isotropic spacing in all dimensions (each voxel represents a volume of $0.5208^3 mm^3$), rescaled to have dimensions 384×384 in the axial plane. Following [94, 129], we remove the empty slices where no prostate marking appears. Each volume is then padded with zeros to have 96 pixels. We use the central crop of each sample of size $256 \times 256 \times 80$ (rescaled down to $192 \times 192 \times 80$ for more manageable data size) to
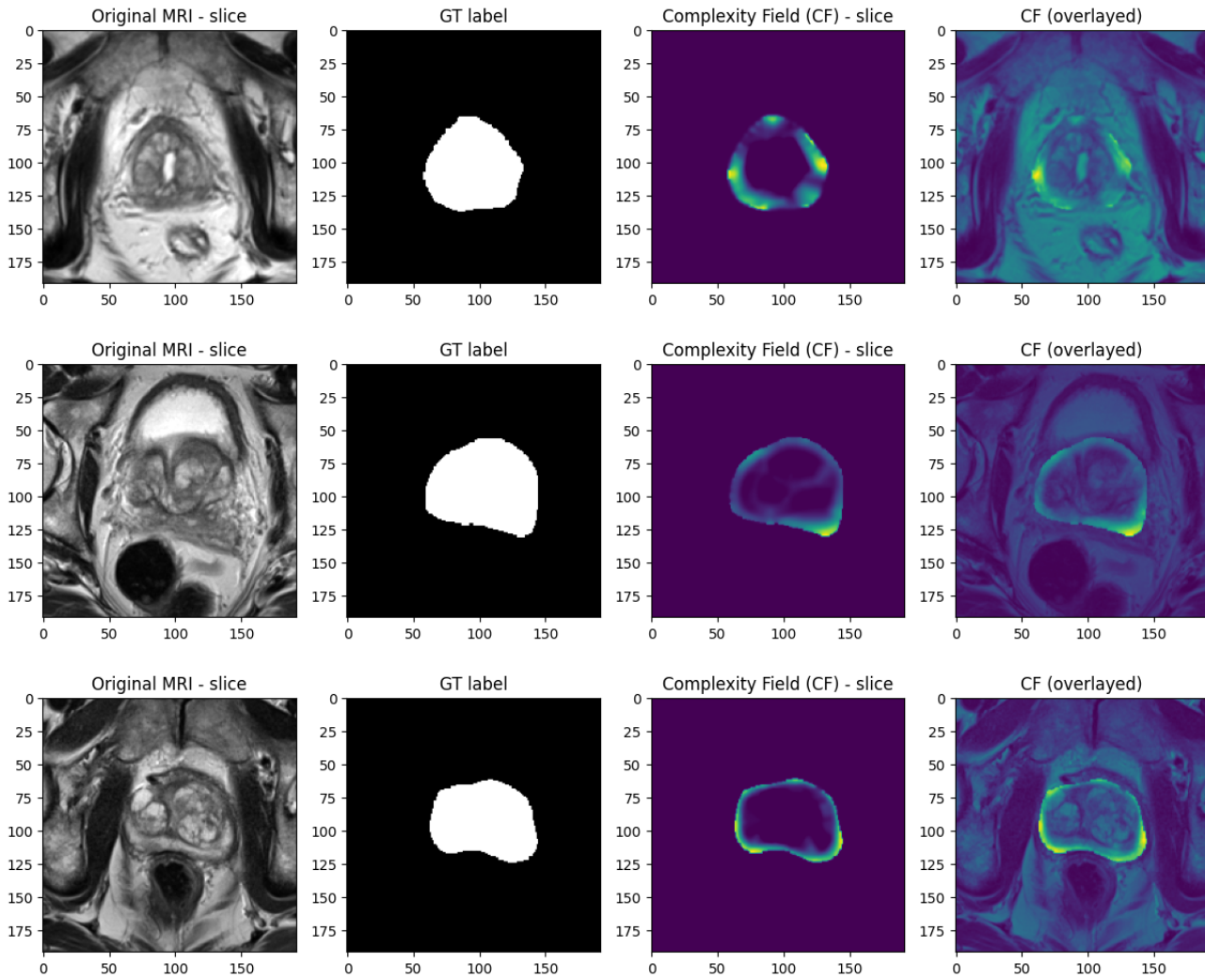
Figure 5.3: Sample slices from MRI data, GT labels, and constructed complexity-encoding fields. The images in the last column are acquired simply as a weighted sum of the complexity fields and the corresponding image slices.

train our networks. Finally, we clip out the top 0.5% of the histograms of each image slice [129].

Augmentations. In [82] researchers have shown that applying deep stacked transformations constitute a strong baseline for generalizable deep learning models in medical image segmentation tasks. Following their approach, during training, we apply the following set of augmentations both for the baseline and our proposed method:

1. Randomized translations in three dimensions where maximum displacements are $(30, 30, 10)$ voxels in respective directions ($p = 0.95$)

2. Randomized rotations in the axial plane where the rotation range is between -30 and 30 degrees ($p = 0.95$)

3. Randomized scaling in the axial coordinates where the scaling range is between 0.8 and 1.2 ($p = 0.95$)

4. Randomized elastic deformations where $sigma$ is in range $[5, 7]$, and the magnitude is in range $[10, 50]$ ($p = 0.5$)

5. Randomized Gaussian noise addition where $\sigma = 0.5$ ($p = 0.5$)

6. Randomized contrast adjustments where γ is in range $[0.5, 4.5]$ ($p = 0.5$)

Model. We employ 3D UNet [148, 149] enhanced with residual units [150] using the implementation of MONAI [151, 152] with parameters $channels = (16, 32, 64, 128, 256)$, number of residual units set to 2, instance normalization as the normalization method, and the number of output channels set to 2; one for the background and one for the target organ.

Model training. We train the networks for both baseline and the proposed method for 1000 epochs with batch size set to 4. Adam optimizer is used with the initial learning rate set to $1e - 3$ and weight decay to $1e - 4$. For learning rate scheduling, we use linear warm-up scheduling for 60 iterations (min. $1e - 4$). We further tested whether learning rate scheduling via cosine annealing throughout the training helps baseline and the proposed method. Based on our experiments, the baseline yielded better generalizable models with cosine annealing while the proposed method yielded better results without it. Thus, for fair comparison, we report results using the suiting approach for each.

Since our interest is in single domain generalization, we select one of the sites (Site A–RUNMC) for training, *i.e.*, as our source domain. The dataset consists of MRI scans of 30 patients. Following [129], we split 21 of these for training, 3 for valida-

tion, and 6 for testing. The models trained and selected based on the data from this site are deployed for use in the unseen/target domains.

Model selection. For both the baseline and the proposed approach, we train multiple (five) models. The best performing model on the validation set of the source domain during each training session is selected. Among the selected models, mimicking the practical scenario, we deploy the model that performs the best on the source domain's test set to unseen/target domains.

5.3.3 Compared methods

The methods we compare to, *i.e.* ADS [86] and BayeSeg [129], use adversarial data augmentation and statistical modelling, respectively.

In more detail, ADS employs adversarial training to enrich the training data to guarantee sufficient coverage of target domains. To this end, the method integrates two convolutional branches, each taking the source image and random noise sample as inputs and outputs two randomly synthesized images. The two branches aim at capturing more possible variations of the source images by means of an adversarial loss: $\ell(\hat{X}_1, \hat{X}_2) = KL(S(\hat{X}_1)\|S(\hat{X}_2))$ where $\hat{X}_1$ and $\hat{X}_2$ are the outputs of the domain synthesizers for a single image X, S is the segmentation branch and KL denotes the Kullback-Leibler divergence. The domain synthesizers try to maximize this loss whereas the segmentation branch attempts to minimize it. The method integrates an additional mutual information branch in order to ensure that the synthesized images do not alter the semantics of the image. This branch attempts to minimize the discrepancy between the features of the real image X and the synthesized images $\hat{X}_i$.

BayeSeg takes a two-staged approach. The input image y first goes through a network that decomposes the image into its shape x and appearance a components. Both the shape and appearance components are modelled so that each follow a probability distribution through elaborate modelling and $y = x + a$. More specifically, a is assumed to follow a Gaussian distribution with mean m (Gamma distribution) and a covariance ρ (Gamma distribution) while x is modelled to be a spatial-correlated simultaneous

64

autoregressive (SAR) model that depends on the edges of the shape (Gamma distribution) and the segmented region (SAR) that depends on segmentation boundaries and class probabilities. After extracting the shape and appearance components, only the shape component is fed through the segmentation network. Both stages are trained simultaneously, where a variational loss term that regularizes the appearance, shape, and segmentation is used along with cross-entropy loss.

5.4   Results and Discussion

We present our results in Table 5.2.

Our method achieves the same Dice score as the baseline on the source domain. For the target domains, the proposed method drastically improves over the baseline on the BIDMC dataset ($+30.7\%$) and in terms of minimum Dice score on targets ($+29.5\%$). For the HK dataset, which differs from the source dataset mainly in field strength (1.5T versus source's 3T), we observe an increase of 7.7% in Dice score. The improvement over baseline in terms of average on targets is $+7.9\%$ which is compatible with the state-of-the-art BayeSeg (ours: 77.3%, BayeSeg: 77.5%, Δ: 0.2%).

Our method achieves a state-of-the-art performance in terms of the minimum Dice score on target domains (ours: 68.3%, BayeSeg: 62.9%, $\Delta = 5.4\%$). Whereas BayeSeg performs the worst on the BIDMC dataset, our method performs the worst on the HCRUDB dataset. Considering the dataset specifications (*cf.* Table 5.1), we argue that this is likely due to the difference in slice thickness of HCRUDB ($1.25mm$) from RUNMC ($3.6 - 4mm$) which is a more challenging issue for 3D segmentation networks. BayeSeg processes MRI data using 2D slices, therefore, it is more likely to be less sensitive to the slice thickness.

We train one extra model for each of baseline and the proposed method, tracking the performance of the model on BIDMC dataset during training. The results are presented in Figure 5.4.

As can be seen, the model trained with our method performs mostly between the $50\% - 70\%$ range in Dice score on the target domain BIDMC, whereas the baseline

Table 5.2: We train the networks on RUNMC dataset. The best model on the validation set is used on the test set of RUNMC and the target domains.

	RUNMC	BIDMC	HK	UCL	BMC	HCRUDB	Avg. on	Min. on
	(source)			(targets)			targets	targets
Baseline	**89.2**	38.8	74.4	85.2	80.7	67.9	69.4	38.8
ADS [86]	–	–	–	–	–	–	71.4	–
BayeSeg [129]	87.3	62.9	**83.2**	81.2	**80.9**	**79.1**	**77.5**	62.9
Ours	**89.2**	**69.5**	82.1	**86.3**	80.3	68.3	77.3	**68.3**
Change over baseline	+0.0	+30.7	+7.7	+1.1	-0.4	+0.4	+7.9	+29.5
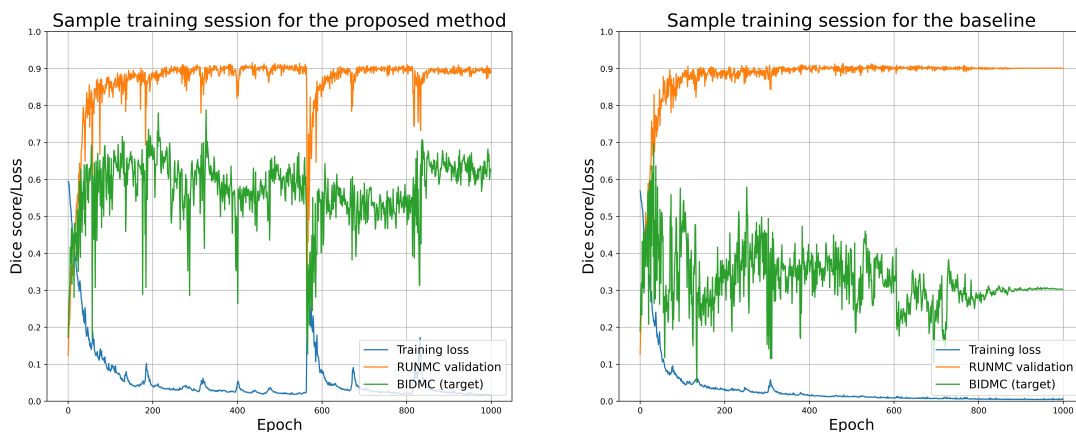
Figure 5.4: Sample training sessions for (left) the proposed method, (right) the baseline where we track the performance of each on one of the target domains (BIDMC).

model oscillates between $20\% - 40\%$ and settles at around 30%. We argue that this is because the baseline model, after the early stages of learning, starts overfitting to the source domain by means of learning ungeneralizable features, whereas the model for proposed method appears to keep on learning generalizable features.

CHAPTER 6

APPLICATION: CURRICULUM LEARNING VIA QUANTIFIED RELATIVE SHAPE COMPLEXITY

Showing that a measure relates to complexity is a non-trivial task due to the lack of a well-defined notion of complexity. Thus, to further demonstrate the relevance of our measure, we employ it on the task of curriculum learning (CL) for neural networks.In this application, we leverage the shape complexity *quantifications* acquired from the different from shape complexity-encoding fields.

In CL, the objective is to plan the training schedule of a neural network according to the complexity or difficulty of examples, aiming to enhance the network's performance. Given the focus of our method on shape complexity, we apply CL on two computer vision tasks: instance segmentation (utilizing the PASCAL VOC 2012 dataset [153]) and image classification (utilizing the CIFAR-10 dataset [154]). Our proposed approach for the image classification scenario is particularly noteworthy, where we explore the application of shape complexity even in the absence of ground truth (GT) object shape masks.

6.1 Related Work

Inspired by how humans and animals learn better when the learning examples are introduced in an organized manner, Bengio et al. [155] introduced CL. As a proof-of-concept, they conduct toy experiments on shape recognition using 2D shapes. They consider the task of classifying shapes into three categories: rectangles, ellipses and triangles. The training data is artificially generated as two different datasets. In one, only regular examples (squares, circles and equilateral triangles) reside, in the other,

the more generic examples of each category. In this setting, they obtain the best generalization when the models are first trained on the dataset with easier examples.

Since its inception, CL has found application across a diverse range of tasks, owing to the idea's independence from specific architectures and applications. In their survey, Wang et al. [156] divides existing methods into predefined CL, where both the sample difficulties and training schedule are fixed and independent of model's learning state, and automatic CL, which is characterized by decisions based on the feedback of the model.

In general, a CL method needs a difficulty measurer, to quantify the easiness of the training data, and a training scheduler, which decides on from what difficulty level should the model learn from throughout the training. A variety of difficulty measurers for predefined CL methods are available. For example, [157] uses data from human experiments, [158] uses image intensity levels, [159] uses number of objects to estimate sample difficulty. Common training schedulers include Baby Step [155, 160], One-Pass [155], root-p [161] schedulers. In Baby Step, the data are divided into clusters based on the difficulty score and the training data, initialized using a small number of these clusters, is gradually expanded while the training progresses. Modifications to Baby Step are considered in [162, 163]. One-Pass is similar, with the difference that chunks used on earlier stages are not utilized in late stages of training. For this reason, One-Pass scheduler is prone to catastrophic forgetting; as a result, is less popular than Baby Step. Continuous scheduling, by means of a pacing function, provides a flexible alternative to Baby Step. In the case of root-p scheduler [161], the chunks with which to expand the training data are determined by the pacing function, given as

$$c_{\text{root}-p}(t) = \min\left(1, \sqrt[p]{c_0^p + (1 - c_0^p) \times t/T_{grow}}\right)$$

where T_{grow} denotes the epoch after which the learner uses whole data, c_0 denotes the initial proportion of easy examples used in training, and $c_{\text{root}-p}(t)$ denotes the proportion of data used at the t-th epoch. Although root-p is found more useful than Baby Step or One-Pass schedulers [161, 164], in general, finding the optimal scheduler depends on the properties of the training data, difficulty measurer and the architecture of the deployed network [156, 165].

The initial studies on CL were focused on learning from the easier samples first. However, further studies [166, 167] report that learning by prioritizing harder examples can also be beneficial. Additionally, [165] proposes two more alternatives: medium-first and two-ends-first. In these alternatives, the easy and hard samples are treated equally whereas the samples with medium difficulty are handled separately.

6.2   In the Presence of Segmentation Labels

In the first scenario for CL, we consider the object detection and instance segmentation problems. We choose PASCAL VOC 2012 [153] as our target dataset. In this dataset there are 20 object classes, including animate objects such as people, cats, dogs, as well as inanimate objects including trains, cars, tables. Each image in the dataset may include multiple instances from these categories. The images are annotated finely in the sense that, for example, occluding objects are removed from the GT label of the relevant object.

6.2.1   Quantification of relative shape complexity

In order to quantify the relative shape complexity of the objects in the dataset, we first extract mean shapes using k-means [168, 169] clustering. We achieve this à la Procrustes: for objects in each category, the convex hulls of GT labels are scaled to have approximately the same area, center-aligned with each other, and, when needed, horizontally flipped to ensure a consistent and standardized representation across different instances. We use convex hulls of GT labels since the occluding objects often interfere with the shape of the object. We avoid applying rotations and vertical flips for further alignment, as we believe these have a larger impact on distorting the recognition of objects [170]. We vectorize the masks and apply k-means clustering with $k = 5$, and use the (de-vectorized) cluster centroids as mean shapes (threshold=0.6). In the first rows of Figures 6.1 and 6.2 such acquired mean shapes can be seen for the train and person categories, respectively.

We acquire shape prototypes by rescaling the mean shapes by $1/15$ in each direction via nearest-neighbor interpolation. With shape prototypes and normalized shape

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
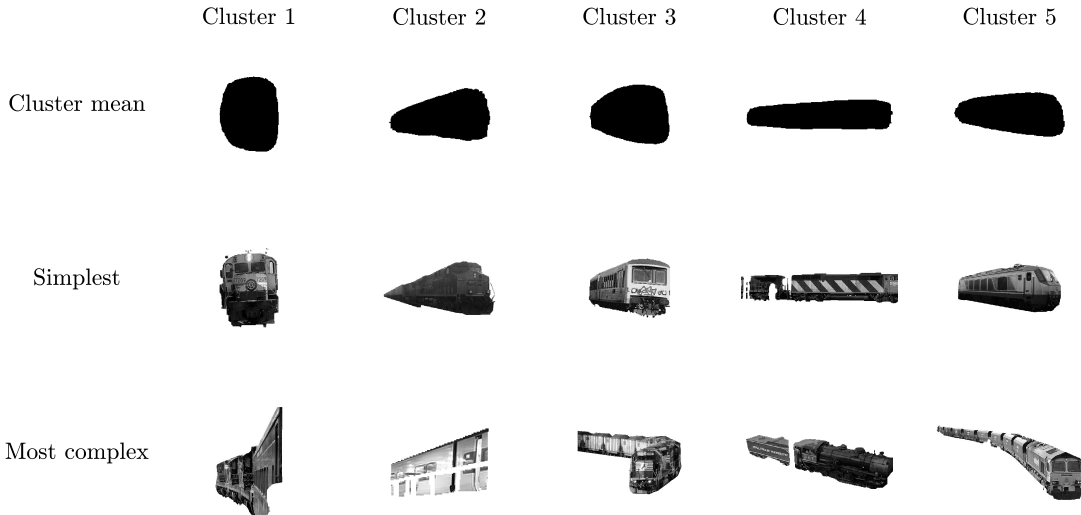
Figure 6.1: Cluster means, the simplest and most complex objects for the train category. When presenting the most complex objects, we exclude those objects with an extremely low pixel count and with masks that are unintelligible.

masks acquired, we construct the complexity-encoding fields over the masks with respect to the corresponding prototypes. Then, we assign each sample a complexity score as described in Chapter 4. Example simplest and most complex objects of each cluster are presented in the second and third rows of Figures 6.1 and 6.2.

Image complexity and image difficulty. We determine the overall complexity of a given image in the dataset by summing the measured complexities of the objects present in that image. It is important to note that this implies images with a greater number of objects are more likely to be considered complex.

To test the feasibility of the image complexity quantifications, we compare our results with a human study conducted by Ionescu et al. [157]. In their study, participants were asked "Is there an {object class} in the next image?". Response times were collected as indicators of image difficulty. Our measurements agree with the collected data on $\approx 65\%$ of image pairs on which image is more complex (averaged over multiple k-means runs and complexity quantifications). We remark that our measurements rely solely on the shapes of objects, discarding image-related information, such as brightness, contrast, texture. For comparison, in the scope of the same study [157], the authors report the average inter-human agreement on a subset of image pairs as
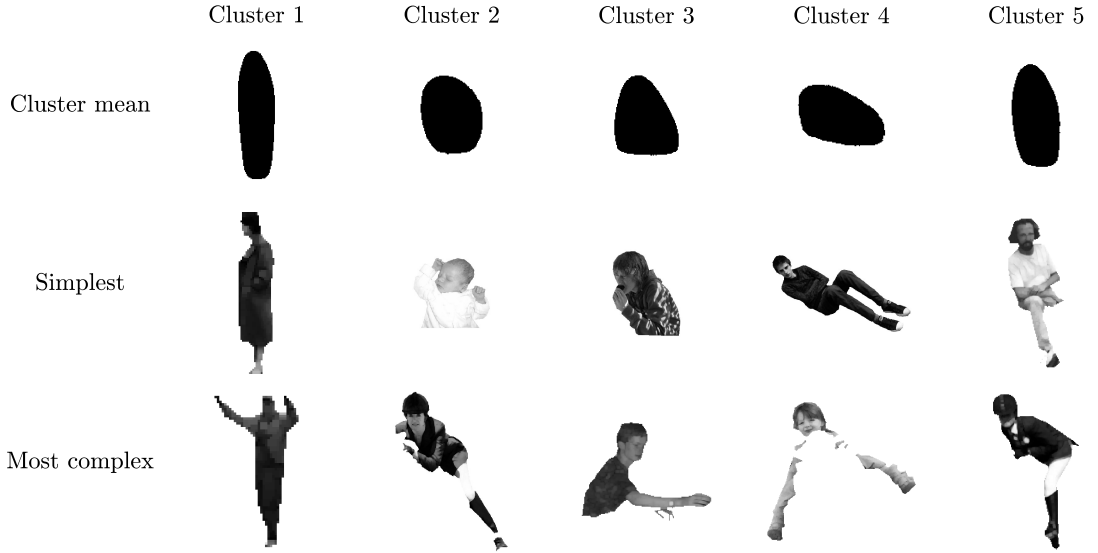
Figure 6.2: Cluster means, the simplest and most complex objects for the person category. When presenting the most complex objects, we exclude those objects with an extremely low pixel count and with masks that are unintelligible.

$\approx 78\%$ and their prediction model (trained on 50% of the collected data, validated on 25%) to agree on $\approx 72\%$ image pairs.

6.2.2 Curriculum learning strategies

We condition the model's curriculum on the training progress and implement curriculum strategies via probabilistic sampling. Our experiments with straightforward implementations of Baby Step, One-Pass, and root-p curricula did not yield convincing results. Thus, we devise two different curricula:

1. from low complexity examples to high complexity examples (L2H)

2. from low complexity examples to all examples (L2A)

For both, the corresponding probability density functions (PDF) use the beta distribution, $f(x;\,\alpha,\beta)$, given as

$$f(x;\,\alpha,\beta) = \frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\text{Beta}(\alpha,\beta)} \quad \text{where} \quad \text{Beta}(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

and Γ denotes the Gamma function.

71

Then, we define the equation of the PDF we use for the L2H as:

$$\mathcal{P}_{\text{L2H}}(x;t) = c(t) \times (f(x; 1 + 6t, 7 - 6t))^{1/2}$$

where $c(t)$ is a normalization function that depends on the parameter t to ensure $\int_0^1 P_{\text{L2H}}(x;t) = 1$. Basically, we keep the sum of the shape parameters α and β as a constant and smoothly pass from right-skewed distributions to left-skewed distributions.

For L2A, we design the PDF equation as:

$$\mathcal{P}_{\text{L2A}}(x;t) = (1 - t)\left(c(t) \times (f(x; 1, 6))^{1/2}\right) + t\,\mathcal{U}(x),$$

where $c(t)$ is as before, and $\mathcal{U}$ is the uniform distribution. In practice, we use the rankings of the samples (divided by the total number) in place of x. In Figure 6.3, we visualize both of the PDFs at different stages of the training.
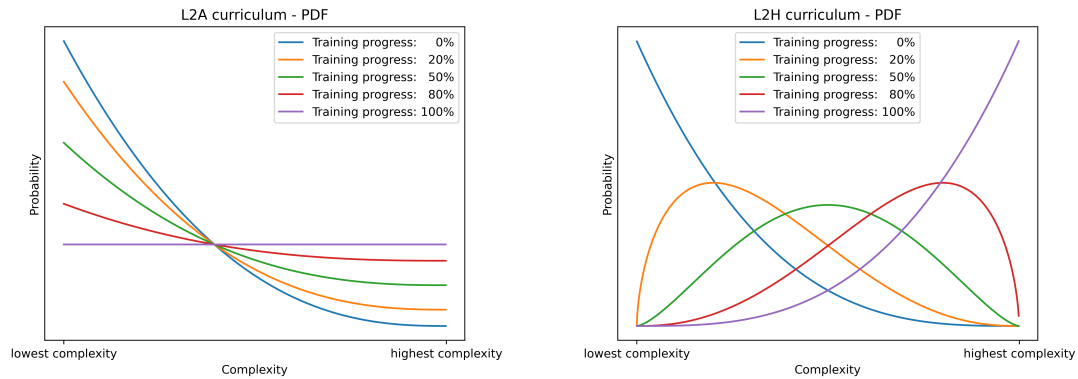
Figure 6.3: The PDFs for low-to-all (L2A, left) and low-to-high (L2H) curriculum strategies at different stages of the training.

6.2.3   Model and training details

In our experiments, we use the PyTorch [171] implementation of the Mask R-CNN [172] model with ResNet-50 [150] backbone. Given the substantial number of parameters in the model, we opt not to train it from scratch; rather, we leverage the pretrained weights from the COCO dataset. We replace the bounding box predictor

of the pretrained network with a randomly initialized Fast R-CNN predictor, and the mask predictor with a randomly initialized Mask R-CNN predictor.

The models are trained on the training split of the PASCAL VOC 2012 dataset for 30 epochs with stochastic gradient descent (momentum set to 0.9, and weight decay set to 10^{-4}). The initial learning rate is 10^{-3}. At epochs 16 and 22 the learning rate is multiplied by 0.1. We use random rotations (rotation angle in range $(-30, 30)$) and horizontal flips for data augmentation.

6.2.4 Results and discussions

We use COCO evaluation metrics to evaluate the models. Following common practice, the maximum detection numbers is set to 100 in all reported results.

In Figure 6.4 we present the average performances of the models on the test set during training. For each method (baseline and two curriculum strategies) we train 5 separate models. We consider both the segmentation and bounding box prediction tasks where both average precision and average recall metrics are used. In the plots, for the sake of better presentation, the results from the initial 4 epochs are omitted.

As evident from the plots, the baseline model, benefiting from fine-tuning on the entire dataset simultaneously, achieves better results on the early epochs. However, its improvement is marginal as the training progresses. In contrast, both curriculum learning methods yield modest results in the initial stages of learning. As the training progresses, so does the learning, and eventually, both curriculum learning methods outperform the baseline.

We separately tabulate the performance of the best models in Table 6.1 and 6.2 since the performance on validation set of each model peaks at a different epoch. We observe that in almost all cases (with the exception of L2H for bounding box prediction with AP@IoU=0.5:0.95 metric), both curricula surpass the baseline. In all cases, the curriculum strategy L2A introduces greater improvements than L2H with lower standard deviations.

PASCAL VOC is an imbalanced dataset. For example, it includes 4,194 instances of
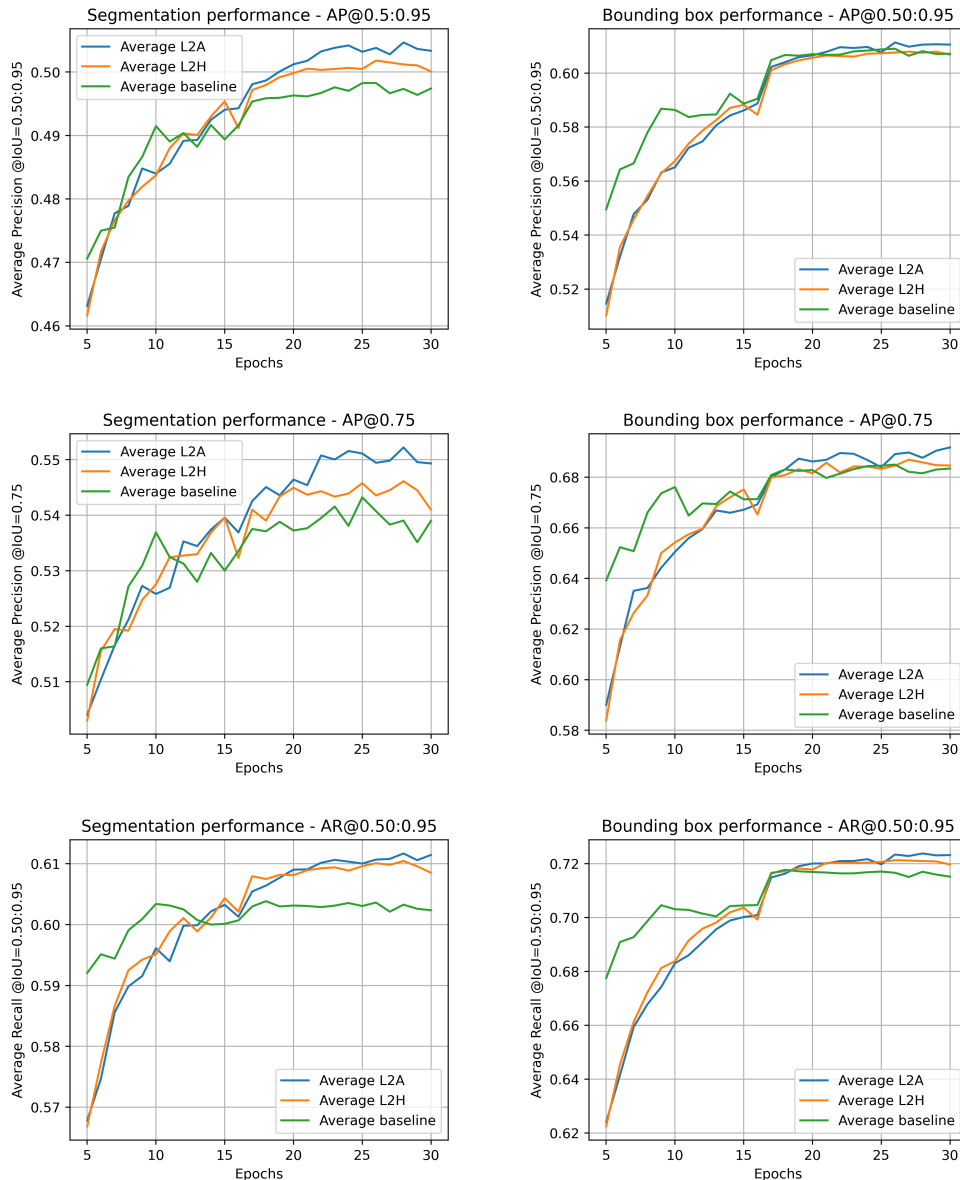
Figure 6.4: Averaged segmentation (left column) and bounding box prediction (right column) performances of the baseline and proposed curriculum learning methods.

the person class spread over 1,994 images in the training set, while only 281 samples from the sofa category (in 257 separate images) exist. Thus, our results are likely improvable by using curriculum learning that also accounts for the imbalance problem. An example of such work is [173]. We did not follow this path because our interest in this chapter is solely to support our claim that the proposed method relates to complexity in the sense of sample difficulty.

Table 6.1: Segmentation performances with respect to Average Precision and Average Recall percentages.

	AP@IoU= $0.5 : 0.95$	AP@IoU= 0.75	AR@IoU= $0.5 : 0.95$
Baseline	49.97 ± 0.10	54.54 ± 0.31	60.66 ± 0.10
L2H	50.35 ± 0.29	55.03 ± 0.49	61.18 ± 0.28
(change)	+0.38	+0.49	+0.52
L2A	$\mathbf{50.60} \pm 0.27$	$\mathbf{55.46} \pm 0.30$	$\mathbf{61.28} \pm 0.19$
(change)	**+0.63**	**+0.98**	**+0.62**

Table 6.2: Bounding box prediction performances with respect to Average Precision and Average Recall percentages.

	AP@IoU= $0.5 : 0.95$	AP@IoU= 0.75	AR@IoU= $0.5 : 0.95$
Baseline	60.97 ± 0.23	68.81 ± 0.31	71.95 ± 0.28
L2H	60.95 ± 0.22	68.93 ± 0.47	72.28 ± 0.19
(change)	-0.02	+0.12	+0.33
L2A	$\mathbf{61.31} \pm 0.32$	$\mathbf{69.41} \pm 0.42$	$\mathbf{72.61} \pm 0.08$
(change)	**+0.34**	**+0.60**	**+0.66**

6.3 When the Segmentation Labels are Missing

Our next scenario is the image classification problem on the CIFAR-10 dataset [154]. The CIFAR-10 dataset is comprised of $60,000$ images of size 32×32 distributed over 10 categories. The dataset lacks GT segmentation labels, thus, we extract approximate shape masks as will be described in § 6.3.1. For each category, we manually provide a set of prototypes, resembling the common shapes pertaining to the category, represented inside a 5×5 grid. After extracting shape masks, we measure the complexity using the prototypes of the corresponding category. We return the minimum value (over different prototypes) as the complexity indicator of an image. The simplest and most complex samples from each category as quantified by our relative shape complexity measure are presented in Figure 6.5.

Figure 6.5: Simplest and most complex samples from each category as quantified by our measure of shape complexity.

6.3.1 Extracting approximate shape masks

As noted, the shape masks are not readily available for the CIFAR-10 dataset. In order to apply our method, we use a modified ResNet-18 architecture [174, 175] where we transpose the problem into instance segmentation. The final layer of ResNet-18 is replaced with a 2D convolutional layer that yields predictions of size $32 \times 32 \times 10$. The class prediction logits are acquired using a smooth-max aggregation over the image domain (32×32).

Using ImageNet1K [176] pretrained ResnNet-18 (the newly introduced layer is initialized with Kaiming normal initialization [177]), we fine-tune the model for 15 epochs. Building on the premise that accurate image classification requires the model to focus on regions occupied by the object itself, we extract pixel-wise confident and correct predictions as a representation of the shape's mask. Note that this is akin to extracting regions where the model attends the most.

76

6.3.2 Curriculum learning strategy

Since our complexity measurements are less precise due to the absence of shape masks, we broadly classify samples into three subsets as low (L), medium (M) and high (H) complexity images. These subsets are approximately equal in size with no overlaps, (1/3rd of all training data in each).

As noted in [156, 165], the correct curriculum strategy depends on the dataset and the difficulty predictor. We acquire results supporting this when we try the curriculum strategies we used in § 6.2 on CIFAR-10 dataset; they do not work. In our experiments, we also find One-Pass scheduler to perform worse than the baseline, possibly due to forgetting the earlier seen samples (catastrophic forgetting). Thus, we cycle through the subsets during training.

Empirically, we find that exposing models to medium complexity samples results in lower accuracies. Therefore, we settle for a curriculum strategy in which both low and high complexity samples are encountered twice as often as the medium complexity ones. This is similar to the "two-ends-first" curriculum of [165].

One cycle of our curriculum strategy involves two passes over L, one pass over M, followed by two passes over H, which we abbreviate as LLMHH. This corresponds to $5/3$ epochs of a network that is trained on all training data.

6.3.3 Model and training details

We use the same modified ResNet-18 model as described in § 6.3.1 with the exception that, this time, we do not use pretrained weights. For the baseline model, we do not use CL, and train the model for 200 epochs. The proposed model is trained for 600 mini-epochs (equivalently, for 120 LLMHH cycles) to ensure that each model learns from the same number of total examples. Both models are trained from scratch with stochastic gradient descent (momentum set to 0.9, and weight decay set to 10^{-4}) with a batch size of 128. We experiment with two different initial learning rates, 10^{-1} (optimal) and 10^{-2}. In both cases, the learning rate is scheduled via cosine annealing (min. 0). The models are trained with binary cross-entropy with logits loss. The used

data augmentations include random crops and spatial transformations.

6.3.4 Results and discussions

We present our results (each averaged over 8 models) in Table 6.3. The accuracy improvements when the curriculum learning is used for both learning rates are statistically significant ($p < 0.01$).

Table 6.3: Averaged accuracy results of the modified ResNet-18 network with and without our proposed curriculum learning strategy for suboptimal ($lr = 10^{-2}$) and optimal ($lr = 10^{-1}$) learning rates.

Method	Initial lr	Accuracy
Modified ResNet-18	10^{-2}	91.10 ± 0.21
+ curriculum (ours)	10^{-2}	$\mathbf{92.19} \pm 0.11$
Modified ResNet-18	10^{-1}	94.60 ± 0.09
+ curriculum (ours)	10^{-1}	$\mathbf{95.03} \pm 0.12$

The improvement in accuracy for the optimal learning rate ($+0.43$) is notably lower than suboptimal case ($+1.19$). One possible reason for this discrepancy could be reaching the capacity limit of the modified network architecture: the best performing (unmodified) ResNet-18 model according to Papers With Code [178], as reported in [179], achieves 95.55% accuracy.

CHAPTER 7

MATHEMATICAL AND NUMERICAL CONSIDERATIONS

7.1 Screened-Poisson Equation and the Laplacian in L^∞

We refer to the equation

$$\left(\Delta_\infty - \frac{1}{\rho^2}\right) f_S = -1 \text{ subject to } f_S\Big|_{\partial S} = 0. \tag{7.1}$$

as the screened-Poisson equation in L^∞ subject to homogeneous Dirichlet boundary conditions due to the replacement of the usual Laplacian with the infinity Laplacian Δ_∞. The infinity Laplacian operator in Euclidean space of arbitrary dimensions is given by

$$\Delta_\infty u = \frac{1}{|\nabla u|^2} \sum_{i,j=1}^n \frac{\partial^2 u}{\partial x_i x_j} \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_j}. \tag{7.2}$$

The operator is the minimizer of $\int |\nabla f|^p$ as $p \to \infty$.

In terms of the usual L^2 Laplacian and the level set mean curvature operator Δ_1, *i.e.*, the 1-Laplacian, $\Delta_\infty = \Delta - \Delta_1$ where

$$\Delta_1 u = |\nabla u|\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right).$$

7.2 Analytical Solutions of the Screened-Poisson Equation in L^∞

Consider the points P_1 and P_2 as given in Figure 7.1. Since they are equidistant from the boundaries, f_S attains the same values at these two points by the above reasoning. Furthermore, this is true for all points having the same y coordinates in the shaded region R_1. In this region, f_S changes in the y direction only, *i.e.* $\partial f_S/\partial x =$

0. Analogous arguments apply for points in R_2 where instead of y, x coordinates determine the equivalence classes. By the continuity of the field on the intersection of R_1 and R_2, the equivalence classes span both regions, and each is a square by itself.
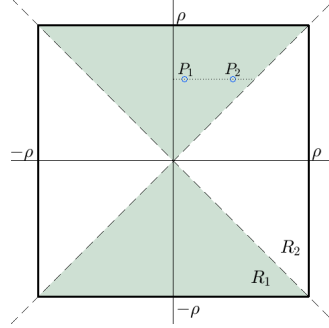
Figure 7.1: Square with sides aligned with grid axes

With these, (7.1) reduces to

$$\frac{\partial^2 f_S}{\partial y^2} - \frac{1}{\rho^2} f_S = -1, \quad \text{for } |y| \geq |x|$$

$$\frac{\partial^2 f_S}{\partial x^2} - \frac{1}{\rho^2} f_S = -1, \quad \text{for } |y| \leq |x| \tag{7.3}$$

$$\text{subject to } f_S\Big|_{\partial S} = 0.$$

In R_1, for the homogeneous part $f_{S,h} = A \exp\{y/\rho\} + B \exp\{-y/\rho\}$, and for the inhomogeneous part $f_{S,p} = \rho^2$. Due to the symmetry of the boundary conditions, the acquired solution is invariant under $y \mapsto -y$ changes. This dictates $A = B$. Applying the boundary condition we acquire

$$f_S\Big|_{R_1} = \rho^2 - \rho^2 \frac{e}{e^2 + 1} \left(\exp\left\{\frac{y}{\rho}\right\} + \exp\left\{-\frac{y}{\rho}\right\} \right).$$

Following the same steps, the solution in R_2 is acquired. The joint solution is given in the closed form as

$$f_S(x, y) = \rho^2 - \rho^2 \frac{e}{e^2 + 1} \times$$
$$\left(\exp\left\{\frac{\max\{|x|, |y|\}}{\rho}\right\} + \exp\left\{\frac{-\max\{|x|, |y|\}}{\rho}\right\} \right). \tag{7.4}$$

80

7.3 Analytical Solutions of the Morphological Laplacian-driven Equation in a Discrete Space

We consider the equation we use in Chapter 4, *i.e.*,

$$\left(\Delta_s - \frac{1}{\rho^2} \right) f_{S,s} = -1, \text{ subject to } f_{S,s}\Big|_{\partial S} = 0, \tag{7.5}$$

where Δ_s is the morphological Laplacian operator with respect to the structuring element (SE) s, as defined in Chapter 4.

In a discrete space we can rewrite (7.5) as

$$\min_{y \in s} f_S(x + y) + \max_{y \in s} f_S(x - y) - \left(2 + \frac{1}{\rho^2} \right) f_S(x) = -1. \tag{7.6}$$

The properties of f_S allow it to be expressed analytically for the family of simplest shapes as shown below. The argumentation is very similar to that in § 7.2, yet, due to the much more generic nature of the neighborhoods, some properties have to be established first.

In the following, we denote $\min_{y \in s} f_S(x + y)$ as m_x and $\max_{y \in s} f_S(x - y)$ as M_x for brevity. We remark that all arguments of locality are with respect to the SE under consideration.

Lemma 7.3.1. $f_S < \rho^2$ on S.

Proof. Assume a point $x' \in S$ exists such that $f_S(x') \geq \rho^2$. Then there must exist a local maximum at a point $\bar{x} \in S$ with $f_S(\bar{x}) \geq \rho^2$. At that point $M_{\bar{x}} \leq f_S(\bar{x})$ and $m_{\bar{x}} \leq f_S(\bar{x})$, and we have

$$M_{\bar{x}} + m_{\bar{x}} - \left(2 + \frac{1}{\rho^2} \right) f_S(\bar{x}) \leq -\frac{f_S(\bar{x})}{\rho^2} \leq -1.$$

The equality holds only when $M_{\bar{x}} = m_{\bar{x}} = f_S(\bar{x}) = \rho^2$. Then, if a point $x \in S$ attains the value ρ^2, by repeated application of the above reasoning, $f_S(R) = \rho^2$ where $R \subseteq S$ is the connected region including x. However, for the points x_b neighboring ∂R, $m_{x_b} = 0$, hence, (7.6) would be violated. $\square$

Lemma 7.3.2. f_S has no local minimum in S.

Proof. Assume f_S has a local minimum at a point $\underline{x} \in S \backslash \partial S$ implying $M_{\underline{x}} \geq f_S(\underline{x})$ and $m_{\underline{x}} \geq f_S(\underline{x})$. But at $\underline{x}$ we have

$$M_{\underline{x}} = f_S(\underline{x}) + \frac{f_S(\underline{x}) - \rho^2}{\rho^2} + f_S(\underline{x}) - m_{\underline{x}} < f_S(\underline{x})$$

where the inequality follows from Lemma 7.3.1. $\qquad \square$

Corollary 7.3.2.1. $f_S > 0$ on S.

Let the notions of a simplest shape, generalized distance transform $t_{S,s}$ with respect to a prototype s, and shape center be as defined in Chapter 4.

Lemma 7.3.3. For a simplest shape S with respect to a SE s, f_S is an increasing function of $t_{S,s}$.

Proof. That f_S is a function of the distance from the boundary follows from the symmetry of the domain with respect to s and the homogeneity of the boundary conditions. To show that f_S is increasing, we note that the contrary would mean a local minimum of f_S exists on S which contradicts with Lemma 7.3.2. $\qquad \square$

Corollary 7.3.3.1. For a simplest shape S, f_S attains its maximum at the shape center.

With these established, we can considerably simplify Eq. (7.6) for a member S of the family of simplest shapes of a prototype s, leading to a second order difference equation, given as

$$\begin{aligned} f_r &= \frac{\rho^2}{2\rho^2 + 1} \left(f_{r-1} + f_{r+1} + 1 \right), \quad 0 < r < \rho \\ f_0 &= 0 \quad \text{and} \quad f_\rho = \frac{\rho^2}{\rho^2 + 1} \left(f_{\rho-1} + 1 \right) \end{aligned} \tag{7.7}$$

where we denote the f's image at the points $\{x \in S : \text{dist}(x, \partial S) = r\}$ as f_r since f_S attains the same value on all members of this set, in view of Lemma 7.3.3. This simplification is very powerful. As shown in Chapter 4, it holds for arbitrary structuring elements of any size, so long as SE itself is connected. The solution to the difference

equation Eq. (7.7) can be obtained as

$$f_r = A\lambda_1^r + B\lambda_2^r + \rho^2,$$

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 - 4\gamma^2}}{2\gamma},$$

$$A = \frac{\rho^2 \left(\lambda_2^\rho - \dfrac{\rho^2}{\rho^2 + 1} \lambda_2^{\rho-1} \right)}{\lambda_1^\rho - \lambda_2^\rho - \dfrac{\rho^2}{\rho^2 + 1} \left(\lambda_1^{\rho-1} - \lambda_2^{\rho-1} \right)},$$

$$B = -A - \rho^2$$

$$(7.8)$$

where $\gamma = \dfrac{\rho^2}{2\rho^2 + 1}$.

The discrete solution Eq. (7.8) is very closely related to the solution acquired in the continuous case Eq. (7.4). To illustrate this, we show both solutions for a diamond in Figure 7.2 where in the continuous case instead of the ℓ^∞ DT we use ℓ^1 DT. We observe that the disagreement between the two fields increase with the distance from boundary. The relative error of the two fields is 4.9×10^{-3}.

| (a) f_1 from Eq. (7.8) | (b) f_2 from Eq. (7.4) | (c) $f_1 - f_2$ |
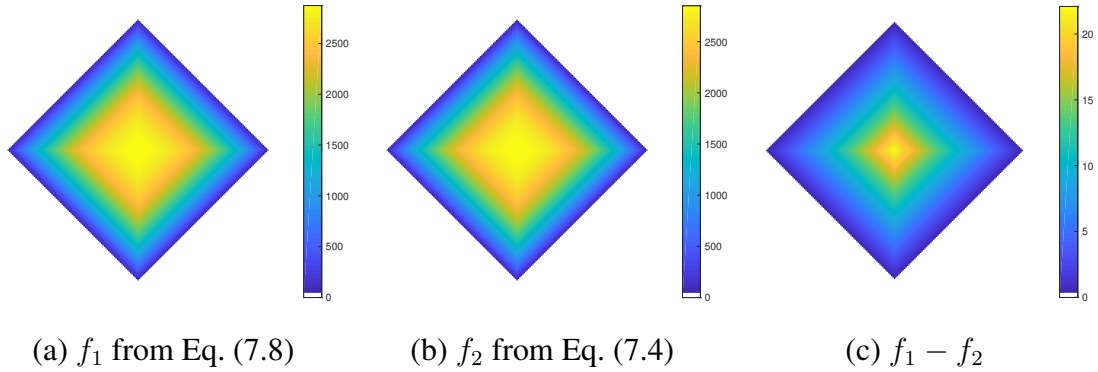
Figure 7.2: Analytical solutions in (a) the discrete space, (b) continuous space, and (c) their difference. Note that, in the depicted fields, the maximum value for both analytical solutions (a) and (b) is around 2,800 whereas at approximately 22 for their difference (c).

In obtaining (7.7) we have assumed that the SE is nonzero at its center, in line with the usual definition of a neighborhood. This choice affects the expression for f_ρ which would otherwise become

$$f_\rho = \frac{\rho^2}{2\rho^2 + 1} \left(2f_{\rho-1} + 1 \right)$$

83

and a solution very similar to (7.8) can be formulated where the only difference is in the expression for the coefficient A which would become

$$A = \frac{\rho^2 \left(\lambda_2^\rho - \frac{2\rho^2}{2\rho^2 + 1} \lambda_2^{\rho-1} \right)}{\lambda_1^\rho - \lambda_2^\rho - \frac{2\rho^2}{2\rho^2 + 1} \left(\lambda_1^{\rho-1} - \lambda_2^{\rho-1} \right)}.$$

7.4 Numerical Solutions

7.4.1 Explicit Euler iteration scheme

We acquire the numerical solutions to Eq. (7.6) via the explicit Euler scheme that follows from the discretization of

$$\partial_t = \left(\Delta_s - \frac{1}{\rho^2} \right) f_{S,s} + 1.$$

The iterative scheme is then given as

$$\begin{aligned}
f_{S,s}^{k+1}(x) = f_{S,s}^k(x) + \\
\Delta k \left(M_x^k + m_x^k - \left(2 + \frac{1}{\rho^2} \right) f_{S,s}^k(x) + 1 \right)
\end{aligned} \tag{7.9}$$

where $f_{S,s}^k(x)$ is the value of the numerically constructed field at x at time k and $f_{S,s}^0 \equiv 0$. The iterations stop whenever maximum point-wise change is less than 10^{-8}.

We identify $\mathcal{F}^k[f_{S,s}](x) = f_{S,s}^k(x) - M_x^k + f_{S,s}^k(x) - m_x^k + \frac{1}{\rho^2} f_{S,s}^k(x) - 1$ in the notation of [180]. Accordingly, $\mathcal{F}$ is an elliptic finite difference scheme, and thus, is monotone and stable. Since $\mathcal{F}$ is strictly increasing in the variable $f_{S,s}^k(x)$ it is also proper. Therefore, the comparison principle holds and solutions are unique. Also, as $\mathcal{F}$ is Lipschitz, the explicit Euler scheme is convergent for $\Delta k > 0$ small enough.

We remark that since the Euler scheme converges to a unique solution, one possible way to speed up the iteration process is to start with $f_{S,s}^0 \equiv f_{S,s}^h$.

The iterative scheme we use do not address the complications arising from the size of the SE. Thus, complexity-encoding fields constructed with respect to SEs of same shape but different size demonstrate what we may refer to as an aliasing effect. An
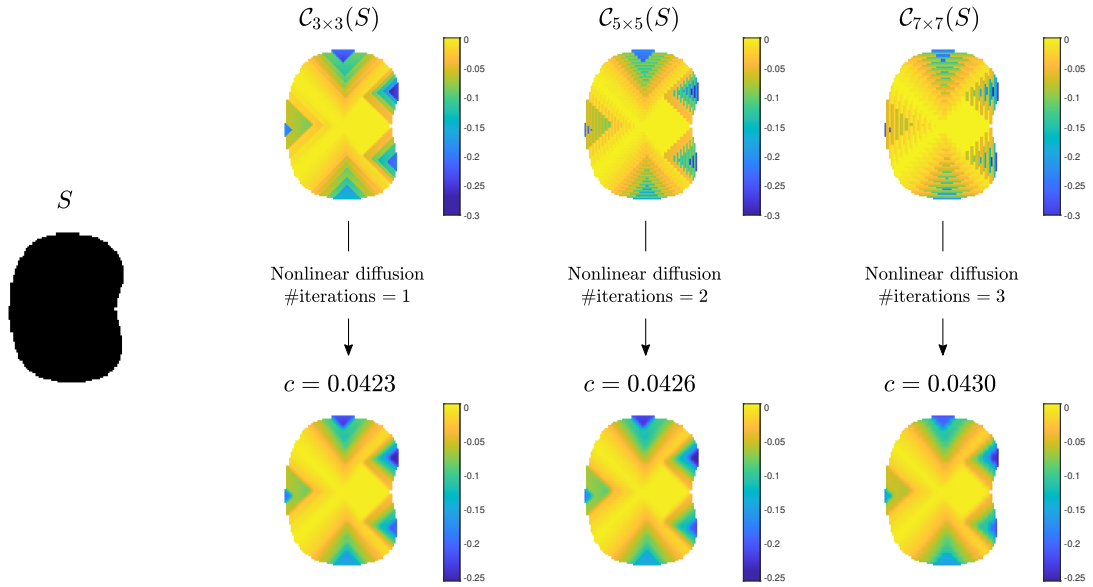
Figure 7.3: Size of the SE affects the complexity-encoding fields (top row). This can be mostly overcome, both in the quantitative and qualitative sense, by diffusion filtering (bottom row).

example is shown on the top row of Figure 7.3. The three fields, from left to right, are acquired using square-shaped SEs encoded in 3×3, 5×5, and 7×7 grids, respectively. This effect does not pose a critical problem; it can be overcome by diffusion filtering. In this case, the diffusion strength is determined by the number of nonzeros of SE.

At the bottom row of Figure 7.3, we visualize the diffused fields. Empirically, we find that nonlinear diffusion with k iterations, where $2k+1$ is the largest dimension of SE, yields both qualitatively and quantitatively plausible results across different SEs.

7.4.2 Interlude: Multigrid methods

Multigrid (MG) methods are iterative methods that use a grid (or mesh) hierarchy to obtain numerical solution to a partial differential equation problem. [181] identifies the two key ideas of MG methods as the smoothing principle (iterations smooth the error) and the coarse grid principle (smoothed errors are well-represented in the coarser grids). The smoothing principle is based on the observation that classical iteration methods smooth high frequency error components efficiently, but fail to do

so for low frequency error components. The second principle relates to the observation that half of the low frequency error components when transferred to a coarser grid become high frequency components. Hence, they can be efficiently removed by the classical iterations on the coarser grid. Taking these two observations together, a recursive strategy can be constructed.

Although the early work in the MG methods were developed for linear problems, they can be used for solving nonlinear problems using the *full-approximation scheme* (FAS, [182]) or *nonlinear multigrid* (NLMG, [183]). The structure of a generic FAS-cycle looks as in Alg. 2. The parameter γ determines the cycling strategy. The case $\gamma = 1$ is referred to as V-cycles and $\gamma = 2$ as W-cycles.

Algorithm 2 Structure of a generic FAS cycle

Input: Initial guess U_0, right hand side f, grid level t, post- and pre-smoothing iteration numbers ν_1, ν_2

Output: Approximate solution U

1: **procedure** FAS(U_0, f, t, ν_1, ν_2)
2:     $U \leftarrow$ Smooth (U_0, f, ν_1) ▷ Pre-smoothing
3:     $r \leftarrow$ Restrict $($Residual $(U))$
4:     $u \leftarrow$ Restrict (U) ▷ Restricted approximation
5:     rhs $\leftarrow r +$ Residual (u)
6:     **if** $t > T$ **then**
7:         $v \leftarrow u$
8:         **for** $\gamma = 1, 2$ **do**
9:             $v \leftarrow$ FAS $(v, \text{rhs}, t-1, \nu_1, \nu_2)$ ▷ Recursive approximation
10:         **end for**
11:     **else**
12:         $v \leftarrow$ solve (u, rhs) ▷ Recursion bottom
13:     **end if**
14:     $E \leftarrow$ Interpolate $(v-u)$
15:     $U \leftarrow U + E$ ▷ Correct the approximation
16:     $U \leftarrow$ Smooth (U_0, f, ν_2) ▷ Post-smoothing
17:     **return** U
18: **end procedure**

When an initial approximation for the field is not present, the *full multigrid* (FMG) can be used together with FAS. In FMG, instead of only correcting the *error* in the initial approximation, the aim is also to get better initial approximations for all the grids in the grid hierarchy. The pseudocode of FMG is presented in Alg. 3.

Algorithm 3 FMG with FAS

Input: Initial guess U_0, right hand side f, grid level t, post- and pre-smoothing iteration numbers ν_1, ν_2 for FMG, post- and pre-smoothing iteration numbers ν_3, ν_4 for FAS

Output: Approximate solution U

1: **procedure** FMFAS(U_0, f, t, ν_1, ν_2, ν_3, ν_4)
2:    **if** $t > T$ **then** rhs
3:       $U \leftarrow$ FMFAS (U_0, f, $t-1$, ν_1, ν_2)   ▷ Recursive approximation
4:    **else**
5:       $v \leftarrow$ solve (u, rhs)   ▷ Recursion bottom
6:    **end if**
7:    $E \leftarrow$ Interpolate ($v - u$)
8:    $U \leftarrow U + E$ ▷ Correct the approximation
9:    $U \leftarrow$ Smooth (U_0, f, ν_2)   ▷ Post-smoothing
10:    **return** U
11: **end procedure**

In general, to obtain an MG method, one should decide at the least on the following:

1. Coarsening of the grids

2. Coarsening of the differential operator

3. Restriction operator

4. Interpolation operator

5. Cycling strategy

6. Smoother

7. Pre-smoothing and post-smoothing iteration numbers

7.4.3 Multigrid implementation and some results

While implementing the multigrid method, we use

1. standard coarsening of the grids (every other pixel is discarded in each direction),

2. the same differential operator for all grids,

3. injection as the restriction operator,

4. both bilinear (for FAS) and bicubic (for FMG) interpolation as interpolation operator,

5. both W-cycles and V-cycles as the cycling strategy,

6. the explicit Euler scheme (§ 7.4.1) for both FMG and FAS cycles,

7. $\nu_1 = 2$, $\nu_2 = 1$ for FAS $\nu_1 = 0$, $\nu_2 = 4$ for FMG with FAS.

In addition to these, we use a scaling factor $s = 1000$ for the restricted defect r (proposed by [183]) which yields a more robust method. These are chosen via exhaustive experimentation with many shapes and structuring elements. The resulting algorithmic procedure to get a solution with arbitrary accuracy is as follows:

Procedure I

1. Apply 1 iteration of FMG with FAS on a uniformly zero field ($\nu_1 = 0, \nu_2 = 4$) to acquire an initial approximation.

2. Apply 50 iterations of FAS with W-cycle ($\nu_1 = 2, \nu_2 = 1$).

3. Apply FAS with V-cycle ($\nu_1 = 2, \nu_2 = 1$) until the desired precision is achieved.

If instead, we are interested in fast approximations, the following algorithmic procedure can be used:

Procedure II

1. Apply 1 iteration of FMG with FAS on a uniformly zero field ($\nu_1 = 0, \nu_2 = 20$) to acquire an initial approximation.

2. Apply 100 iterations of FAS with W-cycle ($\nu_1 = 2, \nu_2 = 1$).

In Figure 7.4 we report results for two selected shapes and the corresponding relative errors of the explicit Euler iterations, Procedure I, and Procedure II. The GT for fields are calculated via explicit Euler iterations that take around an hour and a half each due to the size of the selected shapes. Each method is implemented at the same

optimization level in MATLAB. For the bat figure, Euler scheme (10,000 iterations) took 88 seconds (relative error: 3.8e-2), Procedure I (1 iteration of FMG with FAS, 50 iterations of FAS with W-cycle, 1400 iterations of FAS with V-cycle) took 89 seconds (relative error: **5.8e-5**), Procedure II took **18 seconds** (relative error: 6.1e-3). For the horse figure, Euler scheme took 90 seconds (relative error: 4.5e-2), Procedure I took 88 seconds (relative error: **2.2e-4**), Procedure II took **18 seconds** (relative error: 4.9e-2).
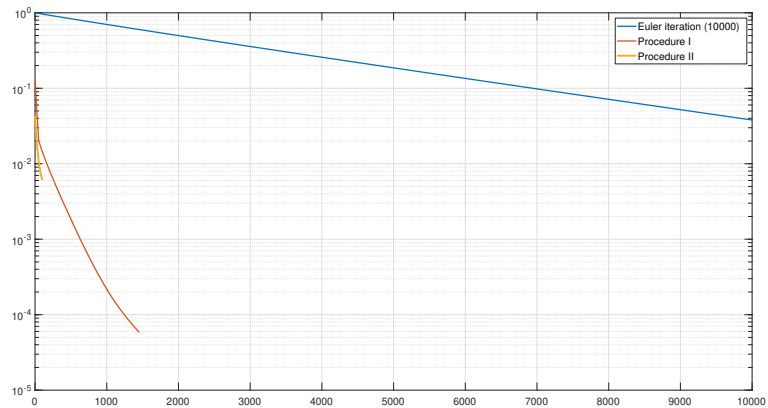
(a) (b)
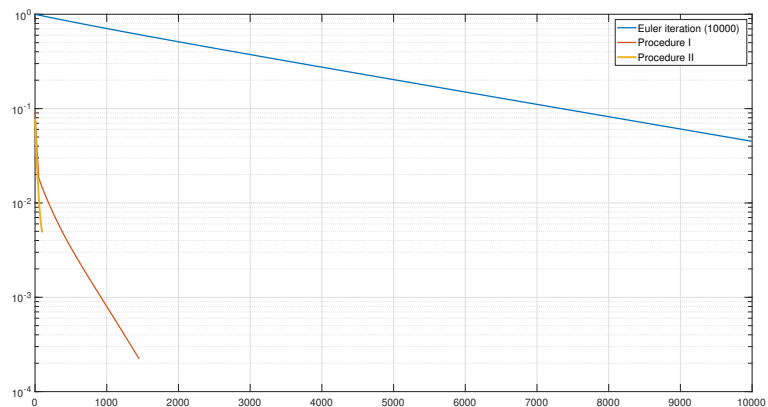
(c) (d)

Figure 7.4: The comparison of relative errors between Euler iterations (10,000), Procedure I, and Procedure II (b)&(d) for the shapes (a)&(c)

CHAPTER 8

SUMMARY

In this thesis, shape complexity is investigated. The considerations included a comparison of shape complexity-related measures on a novel benchmark dataset, proposing a method for quantifying relative shape complexity, and applications to medical image segmentation and computer vision tasks using deep learning.

The proposed benchmark dataset consists of multiple collections of 3D shapes, accounting for different dimensions of shape complexity. Comparing 18 shape complexity measures, we observed that no method excels in all facets. We analyzed the similarities and differences of measures through pairwise correlations and clustering.

The proposed shape complexity measure addresses the phenomenon's relative and emergent aspects. The local and global interaction between a shape's parts are accounted for via a partial differential equation-based formalism. The method encodes the deviation from a simplicity assumption determined by a reference prototype shape. This way, regions that contribute to the overall shape complexity are identified. The proposed method is compatible with arbitrary prototype shapes so long as they have a single connected component. The method is compatible with both continuous and discrete spaces of arbitrary (integer) dimensions. It is comprehensively examined through both empirical investigation and theoretical analysis. Fast numerical solution strategies using multigrid methods are discussed.

In the application of the proposed method to medical imaging segmentation, the single source domain generalization scenario is considered. In this scenario, the models are trained on a single source domain and tested on previously unseen, yet related, target domains which spans a diverse range of variations, including different device

manufacturers, imaging protocols and acquisition conditions.

In the proposed approach, since the convolutional neural networks excel at learning textures, rather than shapes, the focus is on learning generalizable textures. Two main issues prevent this: *i)* inaccurately annotated segmentations may force networks to learn incorrect textures, *ii)* the tissue of a patient may have atypical textures. Both are addressed by leveraging three-dimensional complexity-encoding fields. For each patient, the fields are constructed with respect to the prototype shape which is acquired as the mean shape of the source domain labels. The generalizable texture locations are assumed to correspond to low complexity regions. Thus, the importance of complexity increasing regions are lowered during training. This approach has led to state-of-the-art results on prostate segmentation.

Lastly, through curriculum learning on computer vision tasks, the relevance of the proposed measure to complexity and deep learning applications is further underscored. The image complexities from PASCAL VOC 2012 dataset are predicted via quantifying the complexity of each object in the image. On around 71% of image pairs, the complexity estimations agree with difficulty predictions from human experiments. Curriculum strategies are devised and used in training a Mask R-CNN network. Improved average precision and average recall scores are attained on both bounding box prediction and instance segmentation when the model is trained with the proposed curriculum. In the CIFAR-10 dataset, which is missing segmentation labels, curriculum learning on a modified ResNet-18 model is experimented with. The modified model predicts the class of each pixel in the image. Approximate shape masks are extracted after fine-tuning a pretrained model. Shape complexities are calculated based on these masks. Training the same model from scratch with curriculum learning improved the classification accuracies.

REFERENCES

[1] G. J. Klir, "Complexity: Some general observations," *Systems Research*, vol. 2, no. 2, pp. 131–140, 1985.

[2] M. Gell-Mann, "What is complexity?," in *Complexity and industrial clusters*, pp. 13–24, Springer, 2002.

[3] R. K. Standish, "Concept and definition of complexity," in *Intelligent complex adaptive systems*, pp. 105–124, IGI Global, 2008.

[4] D. Beale, F. Dazzi, and T. Tryfonas, "Unweaving the definitions of complexity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023.

[5] A. Di Ieva, F. J. Esteban, F. Grizzi, W. Klonowski, and M. Martín-Landrove, "Fractals in the neurosciences, part ii: clinical applications and future perspectives," *The Neuroscientist*, vol. 21, no. 1, pp. 30–43, 2015.

[6] K. R. Varshney and A. S. Willsky, "Classification using geometric level sets," *The Journal of Machine Learning Research*, vol. 11, pp. 491–516, 2010.

[7] M. Niimi, H. Noda, and E. Kawaguch, "An image embedding in image by a complexity based region segmentation method," in *Proceedings of International Conference on Image Processing*, vol. 3, pp. 74–77, IEEE, 1997.

[8] Aristotle, "Metaphysics." https://classics.mit.edu/Aristotle/metaphysics.8.viii.html, 350BCE. trans. W. D. Ross.

[9] R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths, and H. Scheingraber, "A comparative classification of complexity measures," *Chaos, Solitons & Fractals*, vol. 4, no. 1, pp. 133–173, 1994.

[10] S. Lloyd, "Measures of complexity: a nonexhaustive list," *IEEE Control Systems Magazine*, vol. 21, no. 4, pp. 7–8, 2001.

[11] M. F. Arslan, A. Haridis, P. L. Rosin, S. Tari, C. Brassey, J. D. Gardiner, A. Genctav, and M. Genctav, "SHREC'21: Quantifying shape complexity," *Computers & Graphics*, vol. 102, pp. 144–153, 2022.

[12] F. Attneave, "Physical determinants of the judged complexity of shapes.," *Journal of experimental Psychology*, vol. 53, no. 4, p. 221, 1957.

[13] H. Thomas, "Spatial models and multidimensional scaling of random shapes," *The American journal of psychology*, vol. 81, no. 4, pp. 551–558, 1968.

[14] M. D. Arnoult, "Prediction of perceptual responses from structural characteristics of the stimulus," *Perceptual and Motor Skills*, vol. 11, no. 3, pp. 261–268, 1960.

[15] W. Saleem, A. Belyaev, D. Wang, and H.-P. Seidel, "On visual complexity of 3d shapes," *Computers & Graphics*, vol. 35, no. 3, pp. 580–585, 2011.

[16] Y. Chen and H. Sundaram, "Estimating complexity of 2d shapes," in *IEEE 7th Workshop on Multimedia Signal Processing*, pp. 1–4, 2005.

[17] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 701–716, 1989.

[18] C. K. Wing, "On the issue of plan shape complexity: plan shape indices revisited," *Construction Management & Economics*, vol. 17, no. 4, pp. 473–482, 1999.

[19] S. Kwon, D. Mun, B. C. Kim, and S. Han, "Feature shape complexity: a new criterion for the simplification of feature-based 3d cad models," *The International Journal of Advanced Manufacturing Technology*, vol. 88, no. 5-8, pp. 1831–1843, 2017.

[20] A. R. Backes, D. M. Eler, R. Minghim, and O. M. Bruno, "Characterizing 3d shapes using fractal dimension," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (I. Bloch and R. M. Cesar, eds.), pp. 14–21, 2010.

[21] K. Arai, "Visualization of 3d object shape complexity with wavelet descriptor and its application to image retrievals," *Journal of Visualization*, vol. 15, no. 2, pp. 155–166, 2012.

[22] S. H. Kim, I. Lyu, V. S. Fonov, C. Vachet, H. C. Hazlett, R. G. Smith, J. Piven, S. R. Dager, R. C. Mckinstry, J. R. Pruett Jr, *et al.*, "Development of cortical shape in the human brain from 6 to 24 months of age via a novel measure of shape complexity," *NeuroImage*, vol. 135, pp. 163–176, 2016.

[23] M. Nitzken, M. Casanova, G. Gimel'farb, A. Elnakib, F. Khalifa, A. Switala, and A. El-Baz, "3d shape analysis of the brain cortex with application to dyslexia," in *IEEE International Conference on Image Processing*, pp. 2657–2660, 2011.

[24] D. Joshi and B. Ravi, "Quantifying the shape complexity of cast parts," *Computer-Aided Design and Applications*, vol. 7, no. 5, pp. 685–700, 2010.

[25] N. Volarevic and P. Cosic, "Shape complexity measure study," *Annals of DAAAM & Proceedings*, pp. 375–377, 2005.

[26] M. Ben Slama, S. Chatti, and B. Louhichi, "Novel method for shape complexity evaluation: a threshold from machining to additive manufacturing in the early design phase," *Research in Engineering Design*, pp. 1–24, 2024.

[27] G. D. Birkhoff, *Aesthetic measure*. Harvard University Press, 1933.

[28] D. L. Page, A. F. Koschan, S. R. Sukumar, B. Roui-Abidi, and M. A. Abidi, "Shape analysis algorithm based on information theory," in *International Conference on Image Processing*, vol. 1, pp. 229–232, 2003.

[29] J. Rigau, M. Feixas, and M. Sbert, "Shape complexity based on mutual information," in *International Conference on Shape Modeling and Applications 2005 (SMI'05)*, pp. 355–360, IEEE, 2005.

[30] M. Genctav, A. Genctav, and S. Tari, "Nonlocal via local—nonlinear via linear: A new distance field via screened Poisson equation," *J. Math. Imaging Vis.*, vol. 55, no. 2, pp. 242–252, 2016.

[31] A. Genctav and S. Tari, "Discrepancy: Local/global shape characterization with a roundness bias," *J. Math. Imaging Vis.*, vol. 61, pp. 160–171, 2019.

[32] M. F. Arslan and S. Tari, "Complexity of shapes embedded in $\mathbb{Z}^n$ with a bias towards squares," *IEEE Transactions on Image Processing*, vol. 29, pp. 8870–8879, 2020.

[33] J. Rossignac, "Shape complexity," *The Visual Computer*, vol. 21, no. 12, pp. 985–996, 2005.

[34] T. Matsumoto, K. Sato, Y. Matsuoka, and T. Kato, "Quantification of "complexity" in curved surface shape using total absolute curvature," *Computers & Graphics*, vol. 78, pp. 108–115, 2019.

[35] J. D. Gardiner, J. Behnsen, and C. A. Brassey, "Alpha shapes: Determining 3D shape complexity across morphologically diverse structures," *BMC Evolutionary Biology*, vol. 18, no. 184, pp. 1–16, 2018.

[36] T. Brinkhoff, H.-P. Kriegel, R. Schneider, and A. Braun, "Measuring the complexity of polygonal objects," in *International Workshop on Advances in GIS*, vol. 109–117, 1995.

[37] M. F. Arslan and S. Tari, "Local culprits of shape complexity," in *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 91–99, Springer, 2021.

[38] F. Wang, B. C. Vemuri, M. Rao, and Y. Chen, "Cumulative residual entropy, a new measure of information & its application to image alignment," in *International Conference on Computer Vision*, vol. 2, pp. 548–548, 2003.

[39] B. B. Mandelbrot, "The fractal geometry of nature," *San Francisco*, 2015.

[40] B. Vasselle and G. Giraudon, "2-d digital curve analysis: A regularity measure," in *International Conference on Computer Vision*, pp. 556–561, 1993.

[41] J. Hayward, J. D. Orford, and W. B. Whalley, "Three implementations of fractal analysis of particle outlines," *Computers & Geosciences*, vol. 15, no. 2, pp. 199–207, 1989.

[42] W. Borkowski, "Fractal dimension based features are useful descriptors of leaf complexity and shape," *Canadian Journal of Forest Research*, vol. 29, no. 9, pp. 1301–1310, 1999.

[43] P. L. Rosin, "Classification of pathological shapes using convexity measures," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 570–578, 2009.

[44] J. Žunić and P. L. Rosin, "A new convexity measurement for polygons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 923–934, 2004.

[45] H. Edelsbrunner and E. P. Mücke, "Three-Dimensional Alpha Shapes," *ACM Transactions on Graphics*, vol. 13, no. 1, pp. 43–72, 1994.

[46] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 73:1–73:12, 2009.

[47] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[48] D. Moser, H. G. Zechmeister, C. Plutzar, N. Sauberer, T. Wrbka, and G. Grabherr, "Landscape patch shape complexity as an effective measure for plant species richness in rural landscapes," *Landscape Ecology*, vol. 17, no. 7, pp. 657–669, 2002.

[49] T.-T. Ngo, C. Collet, and V. Mazet, "Automatic rectangular building detection from vhr aerial imagery using shadow and image segmentation," in *IEEE International Conference on Image Processing*, pp. 1483–1487, 2015.

[50] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 701–716, 1989.

[51] P. L. Rosin and J. Zunic, "Measuring squareness and orientation of shapes," *Journal of Mathematical Imaging and Vision*, vol. 39, no. 1, pp. 13–27, 2011.

[52] A. Genctav and S. Tari, "Discrepancy: Local/global shape characterization with a roundness bias," *Journal of Mathematical Imaging and Vision*, vol. 61, no. 1, pp. 160–171, 2019.

[53] A. Fawzi, S. Moosavi-Dezfooli, and P. Frossard, "The robustness of deep networks: A geometrical perspective," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 50–62, 2017.

[54] A. Oberman, "A convergent difference scheme for the infinity Laplacian: construction of absolutely minimizing Lipschitz extensions," *Mathematics of Computation*, vol. 74, no. 251, pp. 1217–1230, 2005.

[55] M. Breuß and J. Weickert, "Highly accurate PDE-based morphology for general structuring elements," in *Scale Space and Variational Methods in Computer Vision, Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings* (X. Tai, K. Mørken, M. Lysaker, and K. Lie, eds.), vol. 5567 of *Lecture Notes in Computer Science*, pp. 758–769, Springer, 2009.

[56] F. Kuiken, "Linguistic complexity in second language acquisition," *Linguistics Vanguard*, vol. 9, no. s1, pp. 83–93, 2022.

[57] K. Ehret and B. Szmrecsanyi, "Compressing learner language: An information-theoretic measure of complexity in sla production data," *Second Language Research*, vol. 35, no. 1, pp. 23–45, 2019.

[58] F. Van der Slik, R. v. Hout, and J. Schepens, "The role of morphological complexity in predicting the learnability of an additional language: The case of la (additional language) dutch," *Second Language Research*, vol. 35, no. 1, pp. 47–70, 2019.

[59] X. Lu and H. Ai, "Syntactic complexity in college-level english writing: Differences among writers with diverse l1 backgrounds," *Journal of second language writing*, vol. 29, pp. 16–27, 2015.

[60] M. Sipser, "Introduction to the theory of computation," *ACM Sigact News*, vol. 27, no. 1, pp. 27–29, 1996.

[61] H. R. Lewis and C. H. Papadimitriou, "Elements of the theory of computation," *ACM SIGACT News*, vol. 29, no. 3, pp. 62–78, 1998.

[62] D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 439, no. 1907, pp. 553–558, 1992.

[63] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, "Quantum algorithms revisited," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1969, pp. 339–354, 1998.

[64] A. N. Kolmogorov, "On tables of random numbers," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 369–376, 1963.

[65] G. J. Chaitin, "Randomness and mathematical proof," *Scientific American*, vol. 232, no. 5, pp. 47–53, 1975.

[66] G. Pallotti, "A simple view of linguistic complexity," *Second Language Research*, vol. 31, no. 1, pp. 117–134, 2015.

[67] A. Rosenfeld, "Compact figures in digital pictures," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 2, pp. 221–223, 1974.

[68] M. F. Arslan, "On the complexity of shapes embedded in $\mathbb{Z}^n$," Master's thesis, Middle East Technical University, 2019.

[69] A. Oberman, "A convergent difference scheme for the infinity laplacian: construction of absolutely minimizing lipschitz extensions," *Mathematics of computation*, vol. 74, no. 251, pp. 1217–1230, 2005.

[70] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, vol. 1, pp. 424–429, IEEE, 2000.

[71] G. Matheron and J. Serra, "The birth of mathematical morphology," in *Proc. 6th Intl. Symp. Mathematical Morphology*, pp. 1–16, Sydney, Australia, 2002.

[72] Y. Nakagawa and A. Rosenfeld, "A note on the use of local min and max operations in digital picture processing," tech. rep., Maryland Univ College Park Computer Science Center, 1977.

[73] S. R. Sternberg, "Grayscale morphology," *Computer vision, graphics, and image processing*, vol. 35, no. 3, pp. 333–355, 1986.

[74] J. Serra, *Image Analysis and Mathematical Morphology*. Academic Press, 1982.

[75] H. J. A. M. Heijmans, "Theoretical aspects of gray-level morphology," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 06, pp. 568–582, 1991.

[76] R. W. Brockett and P. Maragos, "Evolution equations for continuous-scale morphological filtering," *IEEE transactions on Signal Processing*, vol. 42, no. 12, pp. 3377–3386, 1994.

[77] S. Beucher, *Segmentation d'images et morphologie mathématique*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 1990.

[78] L. J. Van Vliet, I. T. Young, and G. L. Beckers, "A nonlinear laplace operator as edge detector in noisy images," *Computer vision, graphics, and image processing*, vol. 45, no. 2, pp. 167–195, 1989.

[79] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of visual communication and image representation*, vol. 1, no. 1, pp. 21–46, 1990.

[80] J. R. Casas and L. Torres, "Strong edge features for image coding," in *Mathematical Morphology and Its Applications to Image and Signal Processing*, pp. 443–450, Springer, 1996.

[81] B. Burgeth, M. Welk, C. Feddern, and J. Weickert, "Mathematical morphology on tensor data using the loewner ordering," in *Visualization and Processing of Tensor Fields*, pp. 357–368, Springer, 2006.

[82] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[83] S. Cao, N. Konz, J. Duncan, and M. A. Mazurowski, "Deep learning for

breast mri style transfer with limited training data," *Journal of Digital Imaging*, vol. 36, no. 2, pp. 666–678, 2023.

[84] C. Ouyang, C. Chen, S. Li, Z. Li, C. Qin, W. Bai, and D. Rueckert, "Causality-inspired single-source domain generalization for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 4, pp. 1095–1106, 2022.

[85] Z. Zhang, B. Wang, L. Yao, U. Demir, D. Jha, I. B. Turkbey, B. Gong, and U. Bagci, "Domain generalization with adversarial intensity attack for medical image segmentation," *arXiv preprint arXiv:2304.02720*, 2023.

[86] Y. Xu, S. Xie, M. Reynolds, M. Ragoza, M. Gong, and K. Batmanghelich, "Adversarial consistency for single domain generalization in medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 671–681, Springer, 2022.

[87] O. Nuriel, S. Benaim, and L. Wolf, "Permuted adain: Reducing the bias towards global statistics in image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9482–9491, 2021.

[88] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 1501–1510, 2017.

[89] J. Chen, M. Jiang, Q. Dou, and Q. Chen, "Federated domain generalization for image recognition via cross-client style transfer," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 361–370, 2023.

[90] X. Li, Y. Dai, Y. Ge, J. Liu, Y. Shan, and L.-Y. Duan, "Uncertainty modeling for out-of-distribution generalization," *arXiv preprint arXiv:2202.03958*, 2022.

[91] Q. Xu, R. Zhang, Y. Zhang, Y. Wang, and Q. Tian, "A fourier-based framework for domain generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14383–14392, 2021.

[92] Q. Liu, C. Chen, J. Qin, Q. Dou, and P.-A. Heng, "Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1013–1023, 2021.

[93] Q. Dou, D. Coelho de Castro, K. Kamnitsas, and B. Glocker, "Domain generalization via model-agnostic learning of semantic features," *Advances in neural information processing systems*, vol. 32, 2019.

[94] Q. Liu, Q. Dou, and P.-A. Heng, "Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, pp. 475–485, Springer, 2020.

[95] D. Li, J. Zhang, Y. Yang, C. Liu, Y.-Z. Song, and T. M. Hospedales, "Episodic training for domain generalization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1446–1455, 2019.

[96] S. Zhao, M. Gong, T. Liu, H. Fu, and D. Tao, "Domain generalization via entropy regularization," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16096–16107, 2020.

[97] Y. Shi, J. Seely, P. H. Torr, N. Siddharth, A. Hannun, N. Usunier, and G. Synnaeve, "Gradient matching for domain generalization," *arXiv preprint arXiv:2104.09937*, 2021.

[98] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *International conference on machine learning*, pp. 10–18, PMLR, 2013.

[99] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[100] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE transactions on neural networks*, vol. 22, no. 2, pp. 199–210, 2010.

[101] M. Kim and V. Pavlovic, "Central subspace dimensionality reduction using covariance operators," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 657–670, 2010.

[102] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 7, pp. 1414–1430, 2016.

[103] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pp. 41–48, Ieee, 1999.

[104] M. Gong, K. Zhang, T. Liu, D. Tao, C. Glymour, and B. Schölkopf, "Domain adaptation with conditional transferable components," in *International conference on machine learning*, pp. 2839–2848, PMLR, 2016.

[105] Y. Li, M. Gong, X. Tian, T. Liu, and D. Tao, "Domain generalization via conditional invariant representations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[106] Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao, "Deep domain generalization via conditional invariant adversarial networks," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 624–639, 2018.

[107] S. Aslani, V. Murino, M. Dayan, R. Tam, D. Sona, and G. Hamarneh, "Scanner invariant multiple sclerosis lesion segmentation from mri," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 781–785, IEEE, 2020.

[108] A. Khosla, T. Zhou, T. Malisiewicz, A. A. Efros, and A. Torralba, "Undoing the damage of dataset bias," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I 12*, pp. 158–171, Springer, 2012.

[109] M. Ghifary, W. B. Kleijn, M. Zhang, and D. Balduzzi, "Domain generalization

for object recognition with multi-task autoencoders," in *Proceedings of the IEEE international conference on computer vision*, pp. 2551–2559, 2015.

[110] Z. Ding and Y. Fu, "Deep domain generalization with structured low-rank constraint," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 304–313, 2017.

[111] D. Mahajan, S. Tople, and A. Sharma, "Domain generalization using causal matching," in *International Conference on Machine Learning*, pp. 7313–7324, PMLR, 2021.

[112] H. Guan and M. Liu, "Domain adaptation for medical image analysis: a survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, 2021.

[113] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[114] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3723–3732, 2018.

[115] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

[116] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proceedings of the IEEE international conference on computer vision*, pp. 4068–4076, 2015.

[117] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*, pp. 97–105, PMLR, 2015.

[118] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The*

Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14, pp. 443–450, Springer, 2016.

[119] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2507–2516, 2019.

[120] A. Sahoo, R. Panda, R. Feris, K. Saenko, and A. Das, "Select, label, and mix: Learning discriminative invariant feature representations for partial domain adaptation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4210–4219, 2023.

[121] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1406–1415, 2019.

[122] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3722–3731, 2017.

[123] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2107–2116, 2017.

[124] Y. Li, L. Yuan, and N. Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6936–6945, 2019.

[125] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*, pp. 1989–1998, Pmlr, 2018.

[126] A. Mathur, A. Isopoussu, F. Kawsar, N. B. Berthouze, and N. D. Lane, "Flexadapt: Flexible cycle-consistent adversarial domain adaptation," in *2019*

18th IEEE International Conference On Machine Learning And Applications (ICMLA), pp. 896–901, IEEE, 2019.

[127] C. Yang, X. Guo, Z. Chen, and Y. Yuan, "Source free domain adaptation for medical image segmentation with fourier style mining," *Medical Image Analysis*, vol. 79, p. 102457, 2022.

[128] Q. Liu, C. Chen, Q. Dou, and P.-A. Heng, "Single-domain generalization in medical image segmentation via test-time adaptation from shape dictionary," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 1756–1764, 2022.

[129] S. Gao, H. Zhou, Y. Gao, and X. Zhuang, "Bayeseg: Bayesian modeling for medical image segmentation with interpretable generalizability," *arXiv preprint arXiv:2303.01710*, 2023.

[130] Y. Zhang, S. Tian, M. Liao, G. Hua, W. Zou, and C. Xu, "Learning shape-invariant representation for generalizable semantic segmentation," *IEEE Transactions on Image Processing*, 2023.

[131] C. Chen, X. Yang, H. Dou, R. Huang, X. Huang, X. Wang, C. Duan, S. Li, W. Xue, P. A. Heng, *et al.*, "Bridge segmentation performance gap via evolving shape prior," *IEEE Access*, vol. 8, pp. 173961–173973, 2020.

[132] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness," *arXiv preprint arXiv:1811.12231*, 2018.

[133] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9012–9020, 2019.

[134] S. Bohlender, I. Oksuz, and A. Mukhopadhyay, "A survey on shape-constraint deep learning for medical image segmentation," *IEEE Reviews in Biomedical Engineering*, 2021.

[135] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*, pp. 285–296, PMLR, 2019.

[136] P. Bilic, P. Christ, H. B. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, A. Szeskin, C. Jacobs, G. E. H. Mamani, G. Chartrand, *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, p. 102680, 2023.

[137] L. Li, X. Weng, J. A. Schnabel, and X. Zhuang, "Joint left atrial segmentation and scar quantification based on a dnn with spatial encoding and shape attention," in *Medical Image Computing and Computer Assisted Intervention– MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pp. 118–127, Springer, 2020.

[138] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[139] H. Zhang, J. Zhang, R. Wang, Q. Zhang, S. A. Gauthier, P. Spincemaille, T. D. Nguyen, and Y. Wang, "Geometric loss for deep multiple sclerosis lesion segmentation," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 24–28, IEEE, 2021.

[140] F. Navarro, S. Shit, I. Ezhov, J. Paetzold, A. Gafita, J. C. Peeken, S. E. Combs, and B. H. Menze, "Shape-aware complementary-task learning for multi-organ segmentation," in *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pp. 620–627, Springer, 2019.

[141] Y. Wang, X. Wei, F. Liu, J. Chen, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Deep distance transform for tubular structure segmentation in ct scans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3833–3842, 2020.

[142] S. Dangi, C. A. Linte, and Z. Yaniv, "A distance map regularized cnn for cardiac cine mr image segmentation," *Medical physics*, vol. 46, no. 12, pp. 5637–5651, 2019.

[143] N. Baker and J. H. Elder, "Deep learning models fail to capture the configural nature of human shape perception," *Iscience*, vol. 25, no. 9, 2022.

[144] G. Wang, J. Shapey, W. Li, R. Dorent, A. Dimitriadis, S. Bisdas, I. Paddick, R. Bradford, S. Zhang, S. Ourselin, *et al.*, "Automatic segmentation of vestibular schwannoma from t2-weighted mri by deep spatial attention with hardness-weighted loss," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pp. 264–272, Springer, 2019.

[145] N. Bloch, A. Madabhushi, H. Huisman, J. Freymann, J. Kirby, M. Grauer, A. Enquobahrie, C. Jaffe, L. Clarke, and K. Farahani, "Adversarial consistency for single domain generalization in medical image segmentation," in *The Cancer Imaging Archive*, 2015.

[146] G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: a review," *Computers in biology and medicine*, vol. 60, pp. 8–31, 2015.

[147] G. Litjens, R. Toth, W. Van De Ven, C. Hoeks, S. Kerkstra, B. Van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, *et al.*, "Evaluation of prostate segmentation algorithms for mri: the promise12 challenge," *Medical image analysis*, vol. 18, no. 2, pp. 359–373, 2014.

[148] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[149] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Med-

ical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, pp. 424–432, Springer, 2016.

[150] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[151] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[152] E. Kerfoot, J. Clough, I. Oksuz, J. Lee, A. P. King, and J. A. Schnabel, "Left-ventricle quantification using residual u-net," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*, pp. 371–380, Springer, 2019.

[153] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results." http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[154] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," 2009.

[155] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

[156] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4555–4576, 2021.

[157] R. Tudor Ionescu, B. Alexe, M. Leordeanu, M. Popescu, D. P. Papadopoulos, and V. Ferrari, "How hard can it be? estimating the difficulty of visual search in an image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2157–2166, 2016.

[158] L. Gui, T. Baltrušaitis, and L.-P. Morency, "Curriculum learning for facial expression recognition," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 505–511, IEEE, 2017.

[159] Y. Wei, X. Liang, Y. Chen, X. Shen, M.-M. Cheng, J. Feng, Y. Zhao, and S. Yan, "Stc: A simple to complex framework for weakly-supervised semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 11, pp. 2314–2320, 2016.

[160] V. I. Spitkovsky, H. Alshawi, and D. Jurafsky, "From baby steps to leapfrog: How "less is more" in unsupervised dependency parsing," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 751–759, 2010.

[161] E. A. Platanios, O. Stretcu, G. Neubig, B. Poczos, and T. M. Mitchell, "Competence-based curriculum learning for neural machine translation," *arXiv preprint arXiv:1903.09848*, 2019.

[162] T. Kocmi and O. Bojar, "Curriculum learning and minibatch bucketing in neural machine translation," *arXiv preprint arXiv:1707.09533*, 2017.

[163] X. Zhang, G. Kumar, H. Khayrallah, K. Murray, J. Gwinnup, M. J. Martindale, P. McNamee, K. Duh, and M. Carpuat, "An empirical exploration of curriculum learning for neural machine translation," *arXiv preprint arXiv:1811.00739*, 2018.

[164] G. Penha and C. Hauff, "Curriculum learning strategies for ir: An empirical study on conversation response ranking," in *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42*, pp. 699–713, Springer, 2020.

[165] X. Zhou and O. Wu, "Which samples should be learned first: Easy or hard?," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[166] B. Frénay and M. Verleysen, "Classification in the presence of label noise: a survey," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 5, pp. 845–869, 2013.

[167] C. Santiago, C. Barata, M. Sasdelli, G. Carneiro, and J. C. Nascimento, "Low: Training deep neural networks by learning optimal sample weights," *Pattern Recognition*, vol. 110, p. 107585, 2021.

[168] H. Steinhaus *et al.*, "Sur la division des corps matériels en parties," *Bull. Acad. Polon. Sci*, vol. 1, no. 804, p. 801, 1956.

[169] J. MacQueen *et al.*, "Some methods for classification and analysis of multi-variate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

[170] R. Näsänen, A. Syväjärvi, and J. Rovamo, "Effect of image orientation contents on detection efficiency," *Vision research*, vol. 37, no. 8, pp. 1025–1032, 1997.

[171] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[172] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[173] P. Soviany, "Curriculum learning with diversity for supervised computer vision tasks," *arXiv preprint arXiv:2009.10625*, 2020.

[174] S. Liu, C. Lehman, and G. AlRegib, "Robustness and overfitting behavior of implicit background models," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 3274–3278, IEEE, 2020.

[175] C. Lehman, "Learning to Segment CIFAR-10." `https://charlielehman.github.io/post/weak-segmentation-cifar10/`. Accessed: 2024-01-05.

[176] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[177] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[178] "Papers With Code." `https://paperswithcode.com/`. Accessed: 2024-01-05.

[179] T. Moreau, M. Massias, A. Gramfort, P. Ablin, P.-A. Bannier, B. Charlier, M. Dagréou, T. Dupre la Tour, G. Durif, C. F. Dantas, *et al.*, "Benchopt: Reproducible, efficient and collaborative optimization benchmarks," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25404–25421, 2022.

[180] A. M. Oberman, "Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–jacobi equations and free boundary problems," *SIAM Journal on Numerical Analysis*, vol. 44, no. 2, pp. 879–895, 2006.

[181] U. Trottenberg, C. W. Oosterlee, and A. Schuller, *Multigrid*. Elsevier, 2000.

[182] A. Brandt, "Multi-level adaptive solutions to boundary-value problems," *Mathematics of computation*, vol. 31, no. 138, pp. 333–390, 1977.

[183] W. Hackbusch, *Multi-grid methods and applications*, vol. 4. Springer Science & Business Media, 2013.

APPENDIX A

EXPERIMENTS WITH THE SHAPE COMPLEXITY-ENCODING FIELD

In this appendix, we present our further experimentations with the field proposed in Chapter 3. The experiments pertain to the behavior of the complexity field under a variety of deformations. Similar properties hold for the relative shape complexity measure presented in Chapter 4. We present the results for the former only to avoid repetition.

For the notation in the following material, please see Chapter 3. The "s-score" is calculated as $s = \lfloor 1000 \times \text{mean}|d_S| \rfloor$.

A.1 Scale

We start experimenting with the scale of the shapes. It is known beforehand that d_S is scale invariant. This follows from the scale invariancy of both $\hat{f}_{S,numerical}$ and $\hat{f}_{S,assumed}$. We test how scale invariance manifests in discrete space. Based on this we get a sense of the amount that we can regard as an acceptable deviation in s scores in the following results.

In Fig. A.1 we present two hats, where one is the scaled version of the other. The constructed fields behave similarly and s scores differ by 1, indicating that a deviation in s scores by this amount is acceptable. Mean difference in their s scores is $935 \times 10^{-4} \approx 0.001$.

(a) $s = 26$ (b) $s = 25$
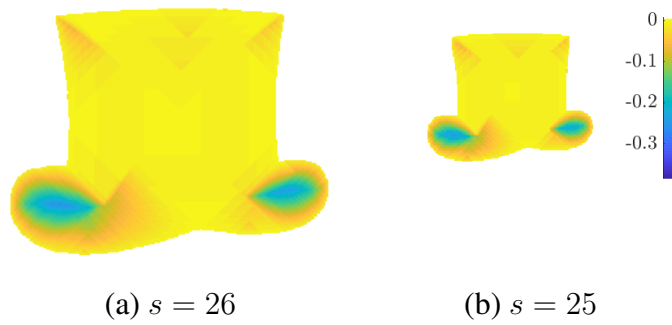
Figure A.1: d_S fields for two hat figures differing in scale.

A.2 Rotation

In L^∞, rotation in the sense of L^2 is not a continuous symmetry. Consequently, d_S is not invariant under rotations.

First, we report results for two pairs of rotated shapes from MPEG7 in Fig A.2. The s scores suggest preserved global similarity of fields under rotation. Yet, they are locally quite different. Observe, for example, the change in d_S around the end of the abdomen of beetle: when rotated, the region becomes aligned with grid axes and leads to significant increase in values of d_S there.

(a) $s = 45$ (b) $s = 46$ (c) $s = 66$ (d) $s = 65$

Figure A.2: s scores are robust under rotation whereas local properties of d_S change for the presented shapes.

To investigate further, we consider eighteen of the shapes presented in this section in Fig. A.3. Although d_S is not continuously invariant under rotations, it is invariant

114

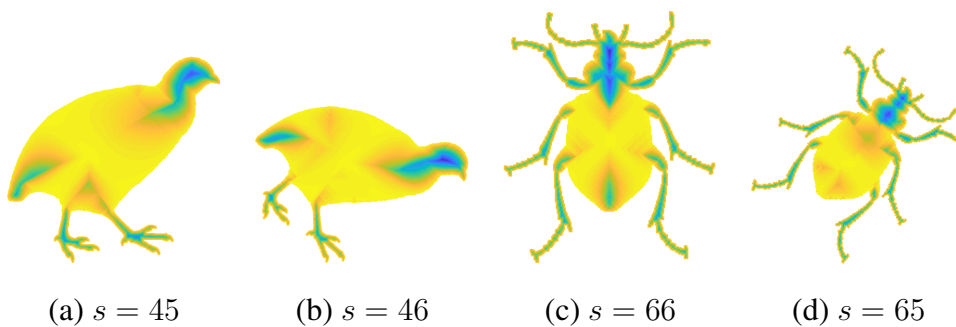under $\pi/2$ rotations, for there is no bias toward x or y directions in its construction. Therefore we rotate each shape by $5k$ degrees for $k \in \{0, 1, \ldots, 18\}$. It is seen that s scores are not stable in general, albeit robust for shapes with circular bodies, such as apples, and shapes with roughly isotropically distributed nonrectangularities, such as beetles. The most varying s scores, in terms of difference in magnitude, are those of shapes acquired by rotating `bat2`, where a difference of 37 is attained between the extrema. Scores from rotated `cup1`, with a maximum of 35 and a minimum of 16, mark the highest extrema ratio. Lastly, remark that rotated shapes acquired from perceptually similar ones go through similar changes in their s scores.
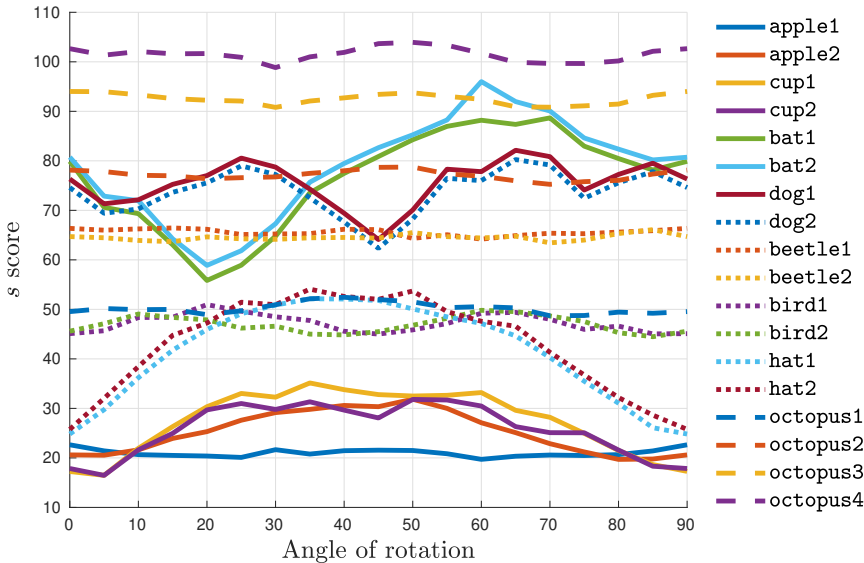
Figure A.3: Plot of s scores versus angle of rotation in degrees. Shapes are enumerated based on their order of appearance in the respective figures.

A.3 Occlusions

More often in real-life applications the acquired shapes are occluded. Therefore, robustness under occlusion for a shape descriptor is desired.

In Fig. A.4 we present an apple, before and after being bitten, where the bite plays the role of occlusion. The ratio of missing pixels in this case to the whole is 7.8%. Qualitatively, we see that the field changes near the bite and is left mostly unaffected elsewhere. To quantify this, we report the percentage of points with a change lesser

than or equal to 0.001, which is 79%. Letting S_1 denote the apple in Fig. A.4(a), and S_2 denote the other, the mean absolute error on S_2 is calculated as

$$MAE = \frac{1}{|S_2|} \int_{S_2} |d_{S_1} - d_{S_2}| = 0.007.$$

Note also that S_2 is assigned a lower score of nonrectangularity as expected since the bite serves to reduce the curviness of the apple.

(a) $s = 23$ (b) $s = 21$

Figure A.4: (a) d_{S_1} and (b) d_{S_2} for an apple before and after being bitten. The bite serves to reduce nonrectangularity as measured by s score.
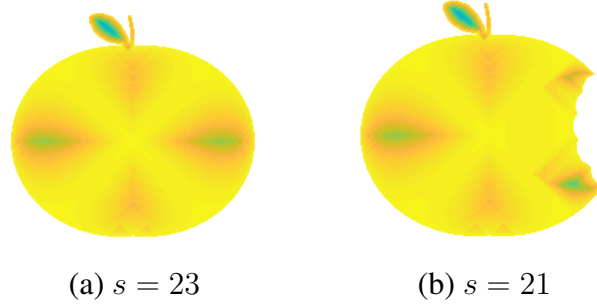
The above robustness-under-occlusion scenario is extended to using eighteen of other shapes. To this end, we synthetically occlude shapes with a circular mask centered near boundary, *i.e.*, at points $t^\infty < 0.2$. This set is further divided into two categories: away from center $g > 0.6$, and near center $g \leq 0.6$, where g denotes the geodesic distance transform calculated from the subset of the shape with $t^\infty \geq 0.6$. In each category, center of occluding circle is chosen randomly. The area of occluding circle, is chosen to be at a ratio of $r = 10\%$, 20% or 30% of the area of shape. For each shape, we create 100 occluded shapes, a sample of which is presented in Fig. A.5.

Percentage of removed points ($R\%$), percentage of points with a change lesser than or equal to 0.001 ($U\%$), mean absolute error, and change in s score (Δs), averaged over 100 shapes is reported in Table A.1. The significant differences between the effects of occlusions away from center and near center justify the division of occlusions into two categories. Position of the occluding mask has a decisive role in MAEs, even disregarding the effective amount occluded. In fact, except for `bat1`, all shapes attain higher MAEs for occlusion near center with $r = 10\%$ than occlusion away from center with $r = 30\%$, despite $R\%$s are higher for the latter. One reason for this is that near center occlusions a have higher chance of changing the shape radius,

116

hence resulting in a change in the governing equation. Another is that introducing new boundaries near the center causes substantial differences in the structure of solutions due to the applied boundary condition and continuity of f_S. It is seen from the results that more convex shapes, such as apples in comparison to beetles, attain higher error rates under occlusions away from center, and vice versa under occlusions near center.

Perceptually similar shapes are observed to attain similar error rates, which, together with the results for rotation, indicate consistency of the method.

(a) away from shape center   (b) near shape center
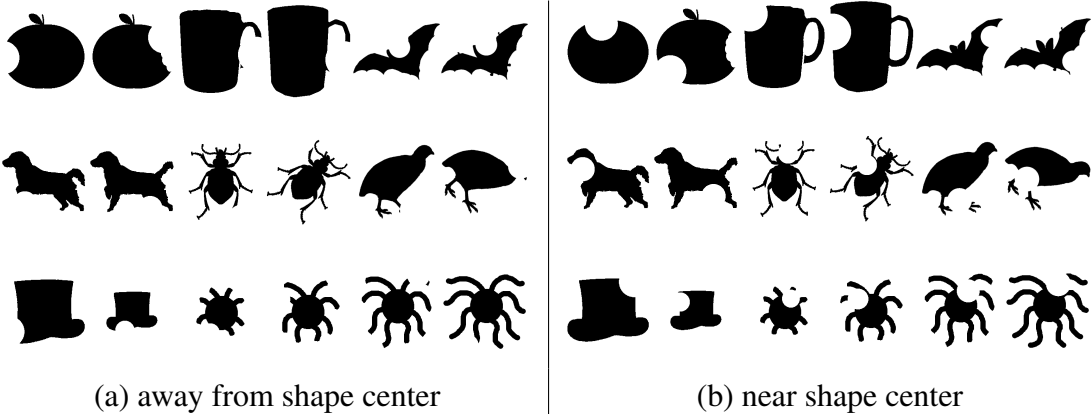
Figure A.5: Sample occluded shapes with $r = 20\%$.

A.4 Topological changes

Tests on occlusion occasionally included topological changes. Here, we deploy two sets to directly experiment with topological changes.

The first set consists of the two cups displayed in Fig. A.6. The discrepancy between the two is at where the handles meet the body and creates a difference in the number of present holes. For both, maximum contribution to nonrectangularity comes from the handles. The nonzero scores attained at bodies of the cups are caused by the curvy bottom and the angled sides. We see that both d_S and s scores are robust in this case.

The second set contains four pocket watches, displayed in Fig. A.7. In the figure, each field is depicted on its own color scale due to the large difference in the s scores between different shapes. We refer to the presented shapes as P_1, P_2, P_3, and P_4, respectively, in the order they are presented in.

117

Table A.1: Percentage of removed pixels ($R\%$), points with a change $\leq 10^{-3}$ ($U\%$), mean absolute error (MAE), and change in s scores (Δs) calculated as averages for 100 occluded shapes acquired by occluding each shape with a disk of area equal to a ratio r of shape area, near boundary and away from, or near shape center.

| Shape | Away from center | | | | | | | | | | | | Near center | | | | | | | | | | | |
| | r = 10% | | | | r = 20% | | | | r = 30% | | | | r = 10% | | | | r = 20% | | | | r = 30% | | | |
	$R\%$	$U\%$	MAE	Δs	$R\%$	$U\%$	MAE	Δs	$R\%$	$U\%$	MAE	Δs	$R\%$	$U\%$	MAE	Δs	$R\%$	$U\%$	MAE	Δs	$R\%$	$U\%$	MAE	Δs
apple1	5	71	.007	.001	9	30	.011	.002	13	20	.014	.005	7	22	.014	.006	12	13	.020	.009	16	7	.025	.014
apple2	5	53	.007	.002	9	24	.013	.005	12	16	.017	.008	7	21	.014	.008	12	15	.019	.011	17	11	.024	.015
cup1	3	76	.002	.002	7	47	.007	.005	9	28	.010	.005	6	32	.015	.011	11	29	.019	.014	16	21	.022	.017
cup2	2	100	.000	.002	4	74	.002	.003	6	39	.005	.003	6	28	.016	.011	11	27	.021	.015	16	19	.027	.020
bat1	4	73	.006	.004	7	27	.015	.008	10	21	.026	.009	6	59	.019	.007	11	25	.026	.007	16	25	.026	.007
bat2	4	86	.003	.003	7	26	.012	.006	11	23	.019	.007	6	52	.020	.006	11	23	.024	.006	16	24	.026	.007
dog1	4	94	.002	.003	6	81	.004	.005	9	71	.008	.006	6	50	.018	.007	11	44	.022	.010	16	42	.025	.010
dog2	4	97	.001	.003	6	82	.004	.005	7	72	.006	.006	6	54	.016	.007	11	50	.021	.010	16	39	.027	.013
beetle1	2	99	.000	.001	4	98	.001	.002	5	96	.002	.003	5	50	.010	.007	10	41	.014	.011	15	23	.022	.015
beetle2	2	99	.000	.001	3	97	.000	.001	5	91	.002	.002	5	54	.010	.006	11	33	.018	.012	15	30	.020	.013
bird1	3	99	.000	.002	4	98	.000	.002	6	93	.002	.003	6	49	.013	.009	11	36	.019	.011	15	23	.026	.017
bird2	4	98	.001	.004	6	94	.002	.005	7	82	.003	.007	6	44	.014	.009	11	28	.021	.013	16	22	.025	.016
hat1	6	67	.006	.005	10	48	.008	.004	13	36	.010	.004	7	36	.016	.015	12	20	.020	.018	18	14	.024	.022
hat2	6	62	.007	.004	10	40	.009	.005	14	29	.012	.005	7	24	.016	.015	12	13	.021	.019	18	10	.024	.020
octopus1	4	93	.001	.003	5	66	.004	.002	8	31	.008	.003	7	24	.012	.007	12	7	.020	.013	16	5	.024	.017
octopus2	4	99	.000	.003	5	97	.001	.004	7	80	.002	.004	6	42	.009	.007	11	14	.017	.013	15	5	.024	.019
octopus3	3	99	.000	.003	5	97	.001	.003	6	94	.001	.005	5	51	.008	.008	9	34	.014	.013	13	18	.018	.016
octopus4	3	99	.000	.002	4	99	.001	.003	6	98	.001	.004	5	56	.007	.007	8	43	.013	.013	13	29	.019	.019

(a) $s = 17$ (b) $s = 18$
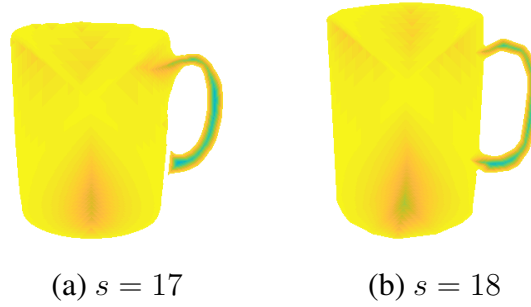
Figure A.6: d_Ss for two cups with different topologies and slight differences in form.

P_1 and P_2 differ in the existing number of holes and boundary detail. Corresponding d_{P_i}s behave similarly in the circular body (*cf.* Fig. A.4) and differ at the bows, where P_1 attains higher scores due to the present hole there. Consequently, P_1 attains a lower, though relatively close, s score. The second two, P_3 and P_4, are significantly different from P_1 and P_2 topologically: P_3 has eleven holes and two different connected components, and P_4 has ten holes and twenty-nine different connected components in the case of 8-connected neighborhoods. This is reflected by the difference in s scores between the two pairs, demonstrating that the positioning of a hole plays a distinctive role as was also observed for the occlusions. For P_3 and P_4, the most rectangular region is found to be crown of the watch. Despite the differences in topological properties, d_{P_3} and d_{P_4} are observed to be similar to each other, and attained s scores are relatively similar.

A.5 Slight changes

We consider changes that can occur between shapes from same categories that are not accounted for in other subsections. To this end, we present three cases each consisting of perceptually similar shapes.

The first case is bats (Fig. A.8(a) and (b)), differing in various disconnected regions, namely, ears, upper wings and lower body. The effect of these differences on s scores and d_S are observed to be negligible. The second case, the two dogs Fig. A.8(c) and (d), have different tails and agree with each other elsewhere. Similar to the first case, d_Ss remain mostly unchanged (except for the tails) and s scores differ only slightly.

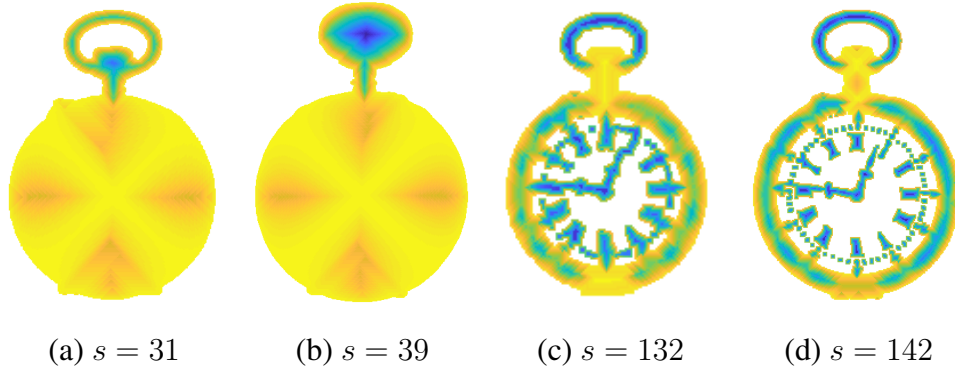(a) $s = 31$ (b) $s = 39$ (c) $s = 132$ (d) $s = 142$

Figure A.7: d_{P_i}s for pocket watches with $H(P_1) = 1$ and $N(P_1) = 1$, $H(P_2) = 0$ and $N(P_2) = 1$, $H(P_3) = 11$ and $N(P_3) = 2$, and $H(P_4) = 10$ and $N(P_4) = 29$, where $H(.)$ denotes the number of holes and $N(.)$ denotes the number of connected components measured by using 8-connectivity.

(a) $s = 78$ (b) $s = 79$ (c) $s = 76$ (d) $s = 75$
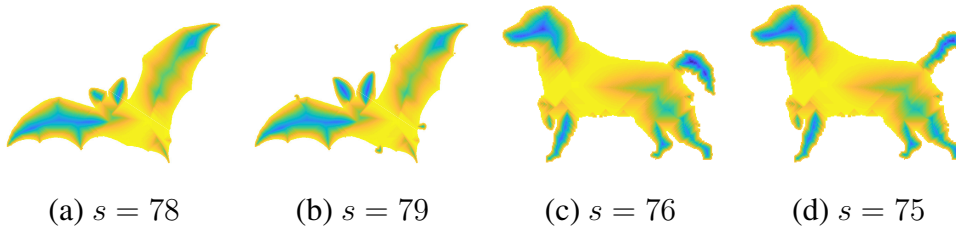
Figure A.8: d_Ss for two pairs of slightly differing shapes.

The last case consists of twenty human face silhouettes differing in pose and around hair, ears, and neck, presented in Fig. A.9. The lowest s score is 26 and the highest is 35 with thirteen of shapes attaining scores between 30 and 32. In all of the shapes, highest contribution toward nonrectangularity is due to necks, whose lengths account for most of the changes in s scores. Ordering by s scores divide the set into two groups based on the poses except for an outlier (fourth shape from left in the second row). The relatively higher s score of the outlier is caused by the introduced deformation in the hair.

A.6   Increasing the nonrectangularity

So far, we tested the robustness of the proposed measure under deformations. Another important aspect of a shape descriptor is to yield consistent results on a sequence of
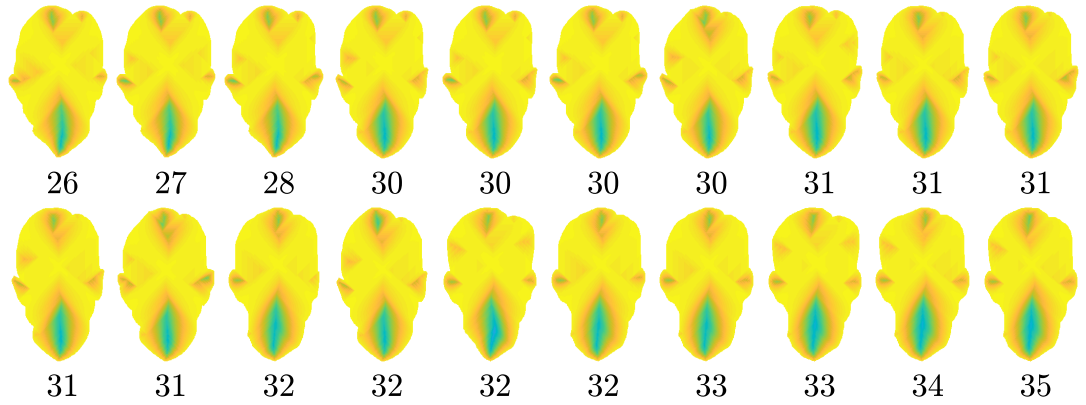
Figure A.9: d_Ss for twenty human faces differing in pose, and around hair, ears, and neck.

shapes for which the order is known beforehand. Here, we consider two sets addressing the consistency of s scores for increasing deviations from rectangularity.

Four octopuses with increasing arm sizes, d_Ss for which are displayed in Fig. A.10, are considered first. The most contribution toward nonrectangularity for all is from the arms. d_Ss at the bodies are the same and very similar to those constructed for other circular bodies. The s scores increase with increasing arms lengths as expected.

The octopuses in MPEG7 dataset consists of five tetrads of octopuses, each tetrad composed of shapes differing in the same vein as the ones in Fig. A.10. For the remaining tetrads the s scores increase with increasing arm lengths too.
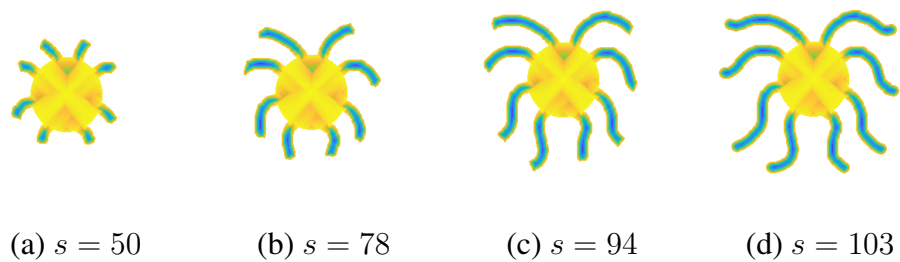

(a) $s = 50$ (b) $s = 78$ (c) $s = 94$ (d) $s = 103$

Figure A.10: d_Ss for four octopuses with differing arm lengths.

Next we consider six square-like shapes (S_1 through S_6) with different forms of deformations, displayed in Fig. A.11. S_1 has a small amount of noise on the boundary and S_2 has a higher amount. The noise is seen to be attenuated quickly and does not

cause a change in the s score of S_1 and affects the s score of S_2 by 1. S_3 and S_4 are both rounded on the edges, the latter is more so, and the former has boundary noise added. Here the rounded edges are observed to cause the structure of fields to become more like those of circular bodies. In this case as well the noise is attenuated near boundary. d_{S_5} is rather interesting: the effect of convex parts is observed to propagate inside whereas that of concave parts does not. Observe that the response of field is proportional to the amount of present deformation. S_6 is considered the most nonrectangular, for which spikes attain the highest scores of nonrectangularity and effect of both spikes and bulges are local.

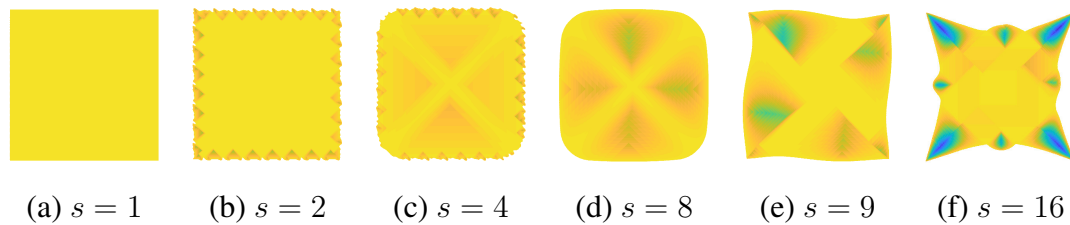The assigned s scores induce a plausible linear order on this set.

(a) $s = 1$ (b) $s = 2$ (c) $s = 4$ (d) $s = 8$ (e) $s = 9$ (f) $s = 16$

Figure A.11: A sequence of square-like shapes with increasing amount of various deformations.

CURRICULUM VITAE

Personal Information

Name, Surname: Mazlum Ferhat, Arslan

Nationality: Turkish

Education

- MS in Computer Engineering, METU, 2019

- BS in Physics, METU, 2017

- BS in Computer Engineering, METU, 2016

Work Experience

- Teaching Assistant, Computer Engineering, METU, 2017 – Present

Publications

- Arslan, M.F., and Tari, S. Measuring relative shape complexity (in preparation)

- Arslan, M. F., Haridis, A., Rosin, P. L., Tari, S., Brassey, C., Gardiner, J. D., Genctav, A. & Genctav, M. (2022). SHREC'21: Quantifying Shape Complexity. Computers & Graphics, 102, 144-153.

- Arslan, M. F., & Tari, S. (2021, April). Local Culprits of Shape Complexity. In International Conference on Scale Space and Variational Methods in Computer Vision (pp. 91-99). Cham: Springer International Publishing.

- Arslan, M. F., & Tari, S. (2020). Complexity of Shapes Embedded in $\mathbb{Z}^n$ with a Bias Towards Squares. IEEE Transactions on Image Processing, 29, 8870-8879.

- Arslan, H., & Arslan, M. F. (2016). The Duffin-Kemmer-Petiau Equation (DKP) Wavefunctions Solutions According to the Virial Theorem for a Spinless Boson Particle Interacting with a Potential $V(r) = kr^n$. Physics Essays, 29(1), 10-14.