

Neural Space-filling Curves

Hanyu Wang, Kamal Gupta, Larry Davis, and Abhinav Shrivastava

University of Maryland, College Park

{hywang66@, kampta@cs., lsd@umiacs., abhinav@cs.}umd.edu

Abstract. We present Neural **S**pace-filling **C**urves (SFCs), a data-driven approach to infer a context-based scan order for a set of images. Linear ordering of pixels forms the basis for many applications such as video scrambling, compression, and auto-regressive models that are used in generative modeling for images. Existing algorithms resort to a fixed scanning algorithm such as Raster scan or Hilbert scan. Instead, our work learns a spatially coherent linear ordering of pixels from the dataset of images using a graph-based neural network. The resulting Neural SFC is optimized for an objective suitable for the downstream task when the image is traversed along with the scan line order. We show the advantage of using Neural SFCs in downstream applications such as image compression. Project page: <https://hywang66.github.io/publication/neuralsfc>.

1 Introduction

In any form of digital communication, information is transmitted via a sequence of discrete symbols. This includes images and videos, even though they are inherently signals with two spatial dimensions (2D). The modus operandi for transmitting such signals is to (1) efficiently encode and quantize their values in the spatial or spectral domain, (2) linearize the signal to a one-dimensional (1D) sequence by using a standard scanning order such as raster, zig-zag, or Hilbert Curve order [11], and finally (3) apply a Shannon [35] style entropy coding technique such as Arithmetic coding [42] or Huffman coding [13] to further compress the 1D sequence. Given the ubiquity of images and videos in our lives, a large amount of effort has gone into optimizing each of these steps of digital communication. The focus of this work is the second step of linearizing the 2D spatial signal to a 1D sequence. A continuous scan order that traverses all spatial locations in two or higher dimensional signals exactly once is also known as the space-filling curve (SFC) [31].

Prior works have proposed various space-filling curves (SFCs) in the last hundred years, most of them context-agnostic, *i.e.*, they are completely defined by the size and dimension of the space without taking into account spatial information of the space, *e.g.*, pixels in the case of two-dimensional images. These universal context-agnostic SFCs are typically defined recursively to ensure simplicity and scale. Some of the SFCs also have spatial coherence properties and have been used in various image-based applications [26,14,2,37].

However, in many applications such as video conferencing, health-care, or social media, the images being transmitted, are often repetitive with a similar

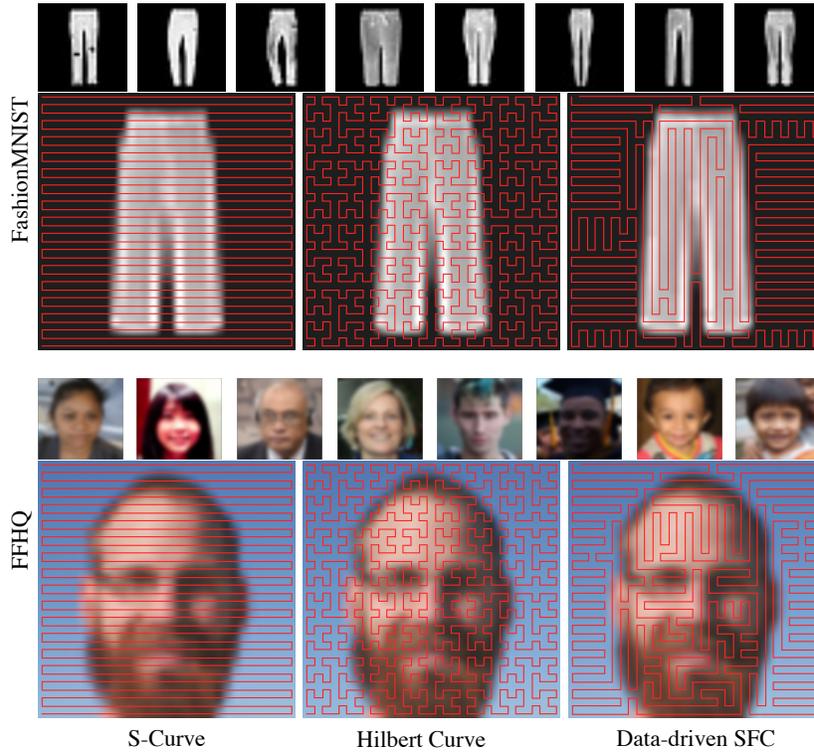


Fig. 1: Given a set of images, a gif, or a video, Neural Space-filling Curves (SFC) can provide a more spatially coherent scan order for images as compared to universal scan orders such as S-curve, or Peano-Hilbert curves. As shown in the example of a trouser and a face, the scan line tends to cover the background before moving to the foreground. (SFCs generated here using half-resolution images and resized for clarity. Best viewed in color.)

layout and content with minor variations transmitted over and over again. GIFs are another great example that consists of highly repetitive content and need to be stored efficiently and often, losslessly. Since universal SFCs do not utilize the intrinsic information of image content, they are far from optimal for a single image or a set of images with repetitive structure (refer to Fig. 1 for an example). Dafner *et al.* [5] proposed SFCs that exploit the inherent correlation between pixel values in an image. Our work improves upon Dafner *et al.*, in two aspects.

- Instead of discovering a single SFC for every image independently, we propose a data-driven technique to find optimal SFCs for a set of images. We postulate that context-based SFCs are more suitable for linearizing a group of images (or a short video/gif), since the cost of storing the SFC itself can be amortized by the number of images.

- We devise a novel alternating minimization technique to train an SFC weights generator, which allows us to optimize for any given objective function, even when not differentiable.

To the best of our knowledge, ours is the first work to propose a machine learning method for computing context-based SFCs and opens new directions for future research on optimal scanning of 2D and 3D grid-based data structures such as images, videos, and voxels. We demonstrate both quantitatively and qualitatively the benefit of our approach in various applications.

2 Related Work

Space-filling Curves (SFCs), introduced by Peano in 1890 [31] and Hilbert in 1891 [11], are injective functions that map a line segment to a continuous curve in the unit square, cube, or hypercube. Most classic SFCs such as Peano-Hilbert, Sierpinski [36], and Moore [27] curves are defined recursively which allows them to scale up to arbitrary resolution with a number of favorable properties. In fact, [21] showed that the entropy of this pixel sequence asymptotically converges to the two-dimensional entropy of the original image for a large number of images coming from sufficiently random sources. Because of their self-organizing capabilities, Hilbert curves have found applications in compression [26,14,2,37], computing [3], recognition [1,20], security [25], databases [18], electronics [46], biology [23] and even web-comics [28]. Hilbert curves have also been used to solve multidimensional task allocation problems in parallel processing [8]. This scheme is used in many job schedulers, such as the famous SLURM [44].

Dafner *et al.* [5] proposed the first context-based SFCs. Their work makes use of cover and merge algorithm [25] to compute SFC for an image. Ouni *et al.* [30] employ image gradient based method to compute context-based SFCs, and they apply their SFCs to lossless compression tasks. Zhou *et al.* [45] improve Dafner *et al.*'s method by considering both data values and location coherency. They also generalize their method to multiscale data via quadrees and octrees.

However, computing a unique SFC specific to an image has limited applications. Compression techniques such as LZW [21] that exploit the correlation between nearby pixel values during encoding, for example, can do a better job with context-based SFCs, however, the cost of storing an SFC itself for each image gives away the advantage. Universal SFCs, such as Hilbert curves don't have this disadvantage and are frequently used in compression applications. In this work, we argue that, applications that need to store and transmit images with repetitive structures such as gifs, can benefit greatly from context-based SFC optimized for that set of images.

Finding an optimal SFC is a combinatorial optimization problem. Solving combinatorial problems have a rich history in the field of machine learning. In 1985, [12] first attempted to solve these problems using a neural network. In the deep learning era, various flavors of attention have been proposed [40,39,7] to reach an approximate solution of NP-hard problems in computer science.

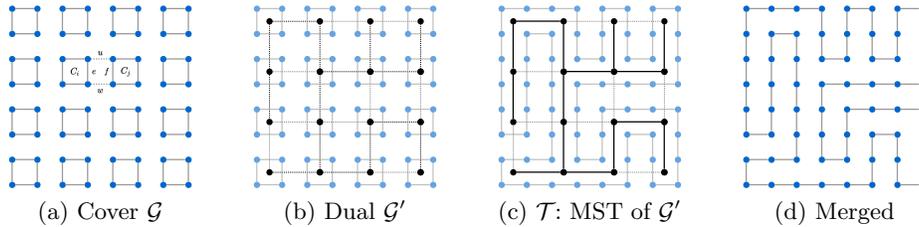


Fig. 2: **Cover and Merge Algorithm.** (a): An 8×8 image fully covered by 2×2 circuits. (b): The dual graph \mathcal{G}' (black) built on the covering circuits \mathcal{G} (blue). (c): The Minimum Spanning Tree \mathcal{T} (black solid lines) of \mathcal{G}' (all black lines) and the Hamiltonian Circuit (blue) induced by \mathcal{T} . (d): A single Hamiltonian circuit merged from the covering circuits. See more details in Sec. 3.1.

Since combinatorial optimization problems are inherently non-differentiable, reinforcement learning (RL) is also a promising alternative for addressing these problems [39,6,7,17]. In our initial experiments, we found the performance of RL approach in our use-case to be unstable and inconsistent.

3 Approach

We first describe the algorithm for computing SFC for one image as proposed by Dafner *et al.* [5] in Section 3.1. We then extend this treatment to a more general setting where we can optimize the SFC for any non-differential objective function for multiple images in Section 3.2. The rest of the Section 3 describes major components of our framework and training procedure in detail.

3.1 Overview of Dafner *et al.* (Single Image-based SFC)

Given an image, [5] represents it as an undirected graph \mathcal{G} whose nodes are pixel locations, and each pixel is connected to its 8 neighboring pixels by an edge. Generating a context-based SFC from the given image is then equivalent to finding a Hamiltonian path in graph \mathcal{G} . They use the cover and merge algorithm (initially proposed by Matias *et al.* [25] to scramble a video signal for secure transmission). As the name suggests, the algorithm works by finding a Hamiltonian path for the image-grid graph \mathcal{G} in two steps - **cover** and **merge**.

In the **cover** step, a dual undirected graph \mathcal{G}' is constructed from \mathcal{G} . The vertices of \mathcal{G}' are small disjoint square circuits covering the whole of \mathcal{G} as shown in Figure 2a. Each circuit covers 4 pixels. We call these initial circuits C_1, C_2, \dots, C_k and connect (C_i, C_j) if the circuits C_i and C_j are adjacent in the original graph \mathcal{G} . Figure 2b shows a dual graph example built for an 8×8 image.

In the **merge** step, all circuits are merged to form a single Hamiltonian circuit. To merge the circuits optimally, a weight is assigned to each edge (C_i, C_j) in \mathcal{G}' representing the “cost” of merging circuits C_i and C_j . The weight $w(C_i, C_j)$

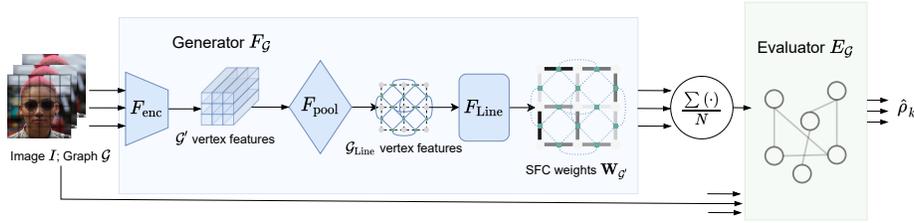


Fig. 3: **Neural SFC pipeline.** The Neural SFC model consists of a weight generator F_G and a weight evaluator E_G . Taking one or more images as input, F_G extracts the deep features of the image and generates the SFC weights. E_G takes both the image(s) and the corresponding SFC weights as input E_G then estimates the negative autocorrelation of the image pixel sequence inferred by the input weights to evaluate the goodness of the input weights with respect to the input image. See more details in Sec. 3.2.

of the edge connecting circuits C_i and C_j is defined as the cost of exchanging the edges e and f with the edges u and w in the image graph:

$$w(C_i, C_j) = |u| + |w| - |e| - |f|, \quad (1)$$

where $|\cdot|$ corresponds to the absolute difference in pixel values at the two vertices of the edge in G . A minimum spanning tree \mathcal{T} is then constructed using these weights. Figure 2c shows what \mathcal{T} might look like for G' . Next, we start merging circuits that are part of \mathcal{T} . Merging two circuits corresponds to removing their adjacent edges in G (e.g., edge e and f in Figure 2a), and creating new edges (e.g., edge u and w in Figure 2a). Note that only adjacent circuit pairs can be merged. Given the spanning tree, all merging operations on G can be done in a linear time to obtain a Hamiltonian circuit as in Figure 2d. Finally, the SFC or Hamiltonian path can be obtained by cutting the circuit at an arbitrary point.

3.2 Neural Space-filling Curves

Dafner’s approach [5] is effective at exploiting the local relationships in an image. However, it has a few limitations.

- Since the edge weights are computed using only two adjacent pixels, the receptive field is limited, and the resulting SFC does not take into account the long-range context in the image.
- Dafner’s approach only works for one image at a time. The notion of context-based SFCs can be further generalized to finding an SFC for a set of images.
- In the current form, [5] cannot optimize for arbitrary objective functions. The context-based SFC obtained is closely tied to edge weights defined by Eq. 1 which encourage autocorrelation with lag-2 in the pixel sequence obtained.

We address each of these issues in our data-driven approach to infer an optimal SFC for a set of images for any objective function. For brevity, we will

optimize for lag- k autocorrelation, however, in our experiments, we will show how we can modify this objective and directly optimize for a downstream application such as compression.

Setup. For the remainder of this section, we use the following notation. We are given a set of N images. Each image (color or grayscale) has a resolution of $H \times W$. For a given image I , we define graph \mathcal{G} over its HW pixel locations. The dual graph \mathcal{G}' consists of 2×2 disjoint circuits covering all the vertices of \mathcal{G} as defined in the previous section. We first note that the whole cover and merge algorithm is context-agnostic barring the step where we assign weights to the edges of the dual graph. Therefore, a context-based SFC can be completely parameterized by these weights, denoted by $\mathbf{W}_{\mathcal{G}'}$. For a given set of weights of image I , the merge operation provides us with a Hamiltonian circuit. A Hamiltonian path, or an SFC, can be obtained by breaking the merged Hamiltonian circuit at any point. Hence, the problem of finding an SFC can be reduced to finding the optimal weights $\mathbf{W}_{\mathcal{G}'}$. We propose to learn a neural network, $F_{\mathcal{G}}$, to approximate $\mathbf{W}_{\mathcal{G}'}$. Once we have the optimized weights, the merge operation is fast and efficient in terms of both memory and speed, and we exploit it to get our desired SFC.

3.3 Weight Generator

The weight generator $F_{\mathcal{G}}$ is designed to take as input a single image I , and output the edge weights $\mathbf{W}_{\mathcal{G}'}$ of its dual graph, *i.e.*, $\mathbf{W}_{\mathcal{G}'} = F_{\mathcal{G}}(I)$. While the input dimension is $H \times W$, the output $\mathbf{W}_{\mathcal{G}'}$, has a size equal to the number of edges in the dual graph \mathcal{G}' . It is trivial to show that for the dual graph \mathcal{G}' consisting of $\frac{H}{2} \times \frac{W}{2}$ vertices (or circuits), the corresponding number of edges will be $\frac{HW-H-W}{2}$. Further, the edges of \mathcal{G}' do not conform to a 2D grid structure as the input image I . To address this, we decompose the weight generator $F_{\mathcal{G}}$ further in three modules.

$$\mathbf{W}_{\mathcal{G}'} = F_{\mathcal{G}}(I) = F_{\text{Line}} \circ F_{\text{pool}} \circ F_{\text{enc}}(I). \quad (2)$$

Figure 3 shows the architecture of the weight generator. The first submodule is a dual graph encoder F_{enc} , which takes I as the input and extracts a deep representation of the vertices of dual graph \mathcal{G}' , resulting in a $\frac{H}{2} \times \frac{W}{2} \times d$ dimensional output, where d is the number of output feature maps. In this work, F_{enc} is implemented using a fully convolutional neural network with residual connections.

The second submodule F_{pool} consists of two pooling filters (and no trainable weights) of dimension 1×2 and 2×1 applied sequentially. F_{pool} imitates the graph operations by aggregating the features of vertices, and hence forming edge features of \mathcal{G}' . Given a d -dimensional representation of edges, we want to compute a scalar weight for each edge in \mathcal{G}' . It is desirable that this scalar weight can exploit not only the edge features, but also long-range relationships among the edges. Hence we construct a Line graph [10] $\mathcal{G}_{\text{Line}}$ to represent the adjacency between edges of \mathcal{G}' . Each vertex in $\mathcal{G}_{\text{Line}}$ corresponds to an edge in \mathcal{G}' . For every

two edges in \mathcal{G}' that have a vertex in common, there is an edge between their corresponding vertices in $\mathcal{G}_{\text{Line}}$.

Using the Line graph, the edge features of \mathcal{G}' becomes the vertex feature of $\mathcal{G}_{\text{Line}}$, and the adjacent relations of edges of \mathcal{G}' are represented by edges of $\mathcal{G}_{\text{Line}}$. To compute the scalar weights on $\mathcal{G}_{\text{Line}}$, we introduce the third submodule, a weights regressor F_{Line} to run on $\mathcal{G}_{\text{Line}}$. F_{Line} can be implemented using Graph Neural Network modules. In this work, we use GCN [16] in MNIST experiments and GAT [38] in FFHQ Faces experiments.

3.4 Objective Functions

The weight generator described above can generate edge weights for a given image. For every mini-batch of images, we take an expected value of the weights for all the images to get $\overline{W}_{\mathcal{G}'}$. Given $\overline{W}_{\mathcal{G}'}$, we can compute an SFC for the mini-batch of images. The quality or ‘goodness’ of an SFC can be different for different applications. In this paper, we consider two plausible objectives: the 1D autocorrelation and the LZW sequence length. For both of them, the first step is to flatten the given image I to the 1D pixel sequence $\{y_i\}$ based on the SFC defined by $\overline{W}_{\mathcal{G}'}$.

Autocorrelation. The 1D autocorrelation measures the internal local similarity of a 1D sequence. Therefore, the smaller the 1D autocorrelation is, the better the SFC is. The lag- k 1D autocorrelation of a pixel sequence $\{y_i\}$ of length HW is defined as

$$\rho_k = \frac{\sum_{i=1}^{HW-k} y_i y_{i+k}}{\sum_{i=1}^{HW} y_i^2}. \quad (3)$$

Code Length. Another SFC quality metric, the LZW sequence length, is inspired from the Lempel-Ziv Welch (LZW) encoding [47,41], which is popularly used to encode GIFs losslessly. Its performance depends on the amount of redundant data in a given sequence. Given a pixel sequence $\{y_i\}$, its LZW length is defined as

$$L = \text{length}(\text{Encode}(\{y_i\})), \quad (4)$$

where Encode is the LZW-encoding function, and length measures the length of a sequence.

Note that computing ρ_k or L requires us to first obtain a minimum spanning tree to infer the SFC from I and $\overline{W}_{\mathcal{G}'}$, which is a non-differentiable operation. Therefore, these metrics, by themselves, cannot be directly used to optimize a neural network. Hence we can’t simply backpropagate gradient information to update $F_{\mathcal{G}}$.

To overcome the problem of non-differentiability of SFC computation, we train an evaluator neural network, $E_{\mathcal{G}}$, as a differentiable proxy to estimate the resulting autocorrelation of SFC weights computed by $F_{\mathcal{G}}$ from the context

image(s). By carefully designing the training procedure, our model E_G is able to approximate any non-differentiable metric, hence serving as an effective loss function of the weight generator. Figure 3 summarizes our approach.

Also note that, while we refer to the lag- k autocorrelation and the LZW length as our objective functions, we can replace them with any loss (or reward) suitable for a given application, even if it is not differentiable.

3.5 Weight Evaluator

The weight evaluator, denoted by E_G , acts as a differentiable proxy to estimate the objective Φ . Depending on the task, we choose $\Phi = -\rho_k$ for minimizing the negative autocorrelation or $\Phi = L$ for minimizing LZW length.

Given the average SFC weights $\overline{\mathbf{W}}_G$, and an image I , we compute

$$\hat{\Phi} = E_G(\overline{\mathbf{W}}_G, I), \quad (5)$$

$$\mathcal{L}_E = \mathbb{E} \left[\|\Phi - \hat{\Phi}\| \right], \quad (6)$$

where \mathcal{L}_E denotes the expected value of l -2 error between groundtruth lag- k autocorrelation and the predicted autocorrelation by E_G computed for the entire batch. \mathcal{L}_E serves as the objective function for training the evaluator.

The inputs to the weight evaluator E_G are $\overline{\mathbf{W}}_G$, and I . Similar to Sec. 3.3, we again decompose the weight evaluator E_G into submodules,

$$\hat{\Phi} = E_G(\overline{\mathbf{W}}_G, I) = E_{\text{Line}}(\overline{\mathbf{W}}_G, \| E_{\text{pool}} \circ E_{\text{enc}}(I)), \quad (7)$$

where $\|$ denotes the concatenation along the feature dimension. Following the graph encoding procedure of the weight generator F_G , E_{pool} and E_{enc} are functionally identical to F_{pool} and F_{enc} , respectively. The final submodule E_{Line} , which takes similar input to F_{Line} , has a different head to predict the estimated objective value $\hat{\Phi}$. The backbone of E_{Line} is also implemented using a GCN or GAT, followed by an average pooling operation and a simple MLP to predict $\hat{\Phi}$ as a single value.

3.6 Training

We adopt an alternating optimization procedure for training the core components of our architecture F_G and E_G . Algorithm 1 gives an overview of the training schema. The weight evaluator E_G solves the regression task of computing the estimated objective $\hat{\Phi}$ for a given set of SFC weights $\overline{\mathbf{W}}_G$. Given the context image I and SFC weights $\overline{\mathbf{W}}_G$, we can get the groundtruth objective by first running Prim’s algorithm [32] to get the desired SFC followed by Eq. 3 or Eq. 4.

Once we have the groundtruth Φ , we optimize a standard L2 loss commonly used in regression methods to train E_G , as we described in Eq. 6. Since we eventually need to use E_G to train the upstream network F , we would like E_G to be trained on a diverse range of input SFC weights \mathbf{W}_G .

The F_G can be trained trivially using a fixed E_G . We empirically observed that training them alternately improves the training dynamics thus boosting the SFC quality.

Algorithm 1: Training Neural SFC

Data: A set of N images each of resolution $H \times W$
Result: SFC weights $\overline{\mathbf{W}}_{\mathcal{G}}$, for the image set

- 1 Randomly initialize $F_{\mathcal{G}}$ and $E_{\mathcal{G}}$;
- 2 **repeat**
 - // training $F_{\mathcal{G}}$*
 - 3 Sample a minibatch of B images;
 - 4 Forward pass for the weight generator $F_{\mathcal{G}}$, $\mathbf{W}_{\mathcal{G}'} \leftarrow F_{\mathcal{G}}(I)$;
 - 5 Expected weights for the mini-batch $\overline{\mathbf{W}}_{\mathcal{G}'} \leftarrow \mathbb{E}(\mathbf{W}_{\mathcal{G}'})$;
 - 6 Forward pass for the weight evaluator $\hat{\Phi} \leftarrow E_{\mathcal{G}}(\overline{\mathbf{W}}_{\mathcal{G}'}, I)$;
 - 7 SGD update for $F_{\mathcal{G}}$ keeping $E_{\mathcal{G}}$ fixed, $\nabla_{F_{\mathcal{G}}} \mathbb{E}[\hat{\Phi}]$;
 - // training $E_{\mathcal{G}}$*
 - 8 For each example in B , with a probability p_1 , get $\mathbf{W}_{\mathcal{G}'}$ using Eq. 1, with a probability p_2 , sample $\mathbf{W}_{\mathcal{G}'} \sim \mathcal{N}(0, 1)$ and with a probability $1 - p_1 - p_2$, keep $\mathbf{W}_{\mathcal{G}'} = F_{\mathcal{G}}(I)$;
 - 9 Run a forward pass for the weight evaluator $\hat{\Phi} \leftarrow E_{\mathcal{G}}(\overline{\mathbf{W}}_{\mathcal{G}'}, I)$;
 - 10 For the whole batch, compute the ground truth Φ using Eq. 3 or Eq. 4;
 - 11 SGD update for $E_{\mathcal{G}}$ with $\nabla_{E_{\mathcal{G}}} \mathcal{L}_E$;
- 12 **until** $\mathcal{L}_E \rightarrow 0$;

4 Experiments

In this section, we evaluate the ability of the proposed training scheme to generate optimized Space-filling Curves (SFC) for a set of images. We further validate the efficacy of the Neural SFCs on real-world applications such as image or gif compression. We compare with standard Raster scan and Hilbert curves. Note that even though Raster scan is not an SFC mathematically speaking, we use it for benchmarking in our experiments due to its prevalence.

4.1 Datasets

We trained the Neural SFC model on four different datasets. Both **MNIST** [19] and **Fashion-MNIST** [43] comprise of 60000 training images, and 10000 test images. Each image is a 28×28 grayscale image, which we resize to 32×32 to do a fair comparison with Hilbert curves which can be defined only when the image resolution is a power of 2. Resizing is done by a simple zero padding around the image. We observe a lot of intra-class similarity in the case of both MNIST and Fashion-MNIST, *i.e.*, the images within the same class are similar in layout and content to each other, and hence we train a single SFC for each MNIST and Fashion-MNIST class.

We also consider **FFHQ** [15] dataset. We downsample all the images to size 32×32 using bilinear interpolation. FFHQ is a dataset of celebrity faces and contains less noise compared to datasets like CelebA [24]. We split the dataset

into 60000 training and 10000 test images. We train a single SFC for all the images in the data.

Lastly, in order to demonstrate a real-world application of SFCs designed for a set of images, we use a large-scale GIF dataset **TGIF** [22]. The dataset consists of 80,000 training gifs, and 11,360 test gifs. We train a single Neural SFC model that takes a gif as an input and outputs an optimized SFC for the gif. We evaluate it on every gif in the test dataset. Average numbers are reported.

4.2 Training details

We consider two different objective functions for training Neural SFC. First, in order to compare our method with Dafner *et al.* [5], we train our model using the autocorrelation objective function. Since Dafner’s method cannot generate SFCs for an image-set trivially, we compute Dafner’s image-set SFC by taking the expected value of \mathbf{W}_G , defined in their method for all the images in the training set. Another possible choice is to calculate an average image for the entire image set, then run Dafner’s method on it to generate an image-set SFC. We use the first setting in all experiments in our main paper because it empirically performs better. In all our experiments, we set $k = 6$, such that lag-6 autocorrelation is used to train the weight evaluator E_G .

Second, we also provide quantitative results on the compression of images and gifs using a Lempel-Ziv encoder (LZW) encoding scheme. Specifically, we optimize the SFC separately for each gif, although, it is possible to train a large SFC encoder on the entire gif dataset. We leave the study for evaluating context-based SFC for large- gifs or video datasets for future work.

4.3 Qualitative Evaluation

Visualizing the 1D sequence. In Figure 4, we show a few examples to illustrate the difference between Neural SFCs and Hilbert Curve on the MNIST, Fashion-MNIST, and FFHQ images. We look at the SFC (red curve overlaid on the image of the digits/clothing/faces). While the Hilbert Curve outputs the same SFC for any 32×32 image in the world, Neural SFC optimizes the SFC for a given set of images. To see this difference, we flatten the SFCs into a 1D sequence (images on the right of the digits/clothing/faces), and observe that Neural SFCs tend to keep better long-range spatial coherence than Hilbert Curve. In both cases, Neural SFCs show pixels in fewer clusters as compared to Hilbert curves. Specifically, Neural SFCs are able to roughly stay in the bright regions until they all get covered. Therefore, bright pixels will mostly gather in one contiguous segment in the 1D sequence inferred from a Neural SFC. In contrast, Hilbert curves often result in multiple clusters of contiguous structures in the 1D sequence. We provide more examples in the supplementary material.

SFCs obtained with different objectives. Fig. 5 shows two different SFCs obtained by our approach when trained with different objective functions. The figure on the left corresponds to SFC optimized for LZW encoding length, while

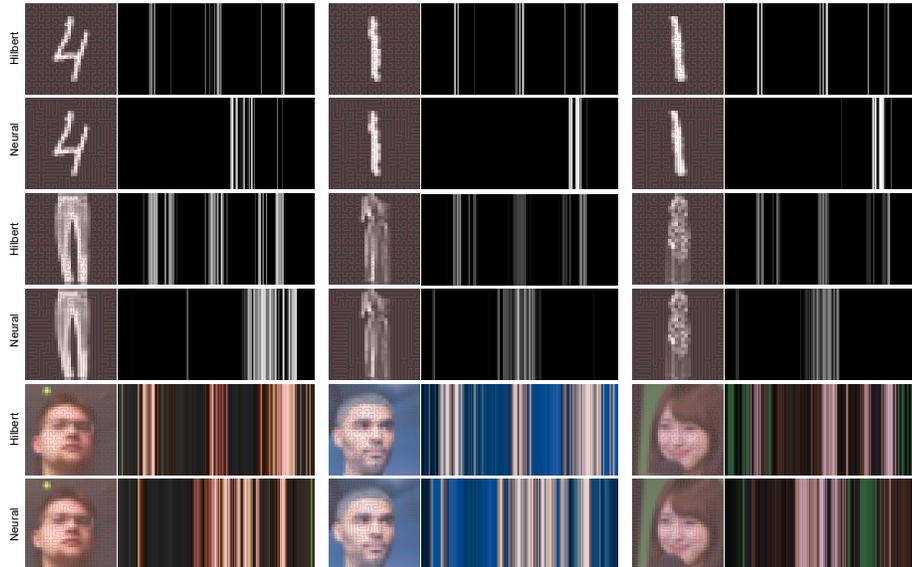


Fig. 4: Qualitative comparison between Hilbert curves and Neural SFCs. Left: SFC (in red color) overlaid on the image. Right: Image flattened according to the SFC and visualized in 1-dimension. Images in the top two rows are from MNIST, the ones in the middle two rows are from Fashion-MNIST, and the ones in the bottom two rows are from FFHQ Faces. Neural SFCs on images from MNIST and Fashion-MNIST are class-conditional, *i.e.*, computed for each class. Therefore, for MNIST and Fashion-MNIST, Neural SFCs in the right two columns are the same since the two images have the same class label. In all datasets, Neural SFCs are more spatially coherent and produce fewer clusters when visualized in 1-dimension. Best viewed in color.

the figure on the right corresponds to SFC optimized for auto-correlation. We observe that generally, SFCs optimized for LZW encoding length are better at short-lag auto-correlation (as compared to SFCs optimized for lag-6 as in most of our experiments). This results in an SFC with fewer turns and straighter paths.

4.4 Optimizing Autocorrelation

To quantitatively evaluate the generated SFCs, we plot lag- k autocorrelations of pixel sequences obtained from test sets of three different datasets MNIST, Fashion-MNIST, and FFHQ. Note that even though, we trained our models only to optimize the lag-6 autocorrelation, we plot them for a range of values of lag- k . From Figure 6, we first observe that the trend is somewhat consistent across multiple datasets, even though they are very different in nature. Neural SFC performs the same or slightly worse than other SFCs at lower values of k . However, for $k > 4$, it outperforms other SFCs by a wider margin. This is

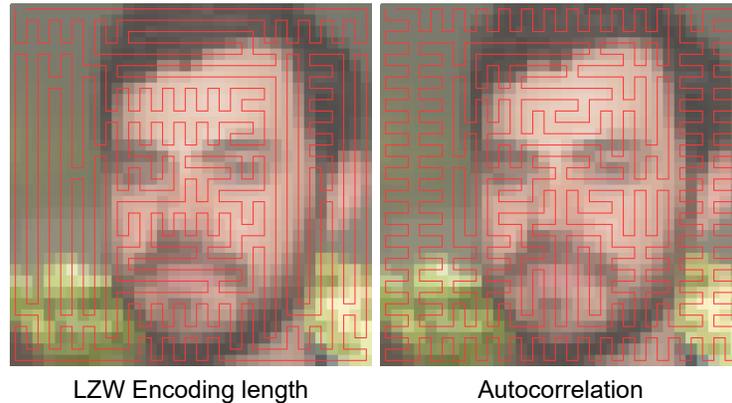


Fig. 5: We visualize the Neural SFCs training with two objectives considered in this paper – autocorrelation and LZW encoding.

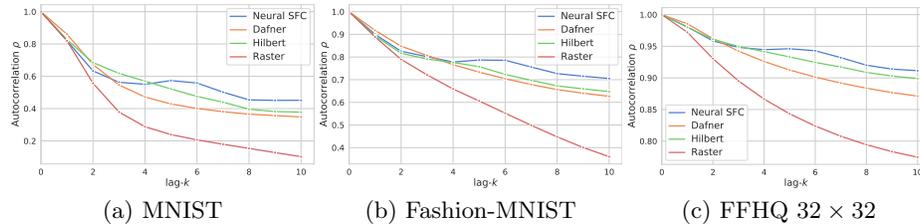


Fig. 6: lag- k autocorrelation for MNIST, Fashion-MNIST and FFHQ datasets. While Dafner *et al.* provide higher autocorrelation for small lag, *i.e.*, from $k = 2$ to $k = 4$, Neural SFCs outperform Dafner *et al.* for $k > 4$ in all the datasets. Note that we trained our model for $k = 6$, and hence this behaviour is expected.

intuitive since our model was optimized to increase the autocorrelation for a large value of lag. This also reflects our model’s ability to capture long-range and global information. We believe that higher gains can be obtained if we train and test using the same k value.

In the second and the third columns of Table 1, we show the autocorrelation values for MNIST, Fashion-MNIST, and FFHQ. We observe that the performance of Dafner’s SFCs is even worse than the Hilbert Curve in most cases. This indicates that the naive average of several good SFCs may not be a good way to compute the SFC for a set of images.

4.5 Optimizing Code Length

In this section, we study how Neural SFC can improve image compression results. Specifically, we use the LZW length objective to optimize $E_{\mathcal{G}}$. This means we set $\Phi = L$ in Algorithm 1, where L is defined in Eq. 4. We evaluate our model’s

Table 1: Comparison of performance of lag- k autocorrelation and LZW Encoding length for different orders. For autocorrelation, we consistently outperform both the universal and context-based SFC computation approaches at high values of k . For LZW Encoding length, we measure the average size per frame in bytes as well as the relative improvement compared to the raster scan order, in the case of each of the datasets. We consistently outperform compression performance for other order schemes.

Dataset	Method	$\rho_6 \uparrow$	$\rho_{10} \uparrow$	Size in bytes (Δ) \downarrow
MNIST	Raster	0.206	0.102	175.4
	Hilbert	0.475	0.378	182.7 (+7.3)
	Dafner [5]	0.401	0.348	-
	Neural SFC (Ours)	0.558	0.451	171.1 (-4.3)
FMNIST	Raster	0.552	0.360	425.8
	Hilbert	0.723	0.647	427.3(+1.5)
	Dafner [5]	0.704	0.627	-
	Neural SFC (Ours)	0.786	0.705	412.4 (-13.4)
FFHQ	Raster	0.824	0.775	688.0
	Hilbert	0.924	0.899	689.6(+1.6)
	Dafner [5]	0.901	0.871	-
	Neural SFC (Ours)	0.943	0.911	678.3 (-9.7)
TGIF	Raster	-	-	563.9
	Hilbert	-	-	567.0 (+3.1)
	Neural SFC (Ours)	-	-	556.9 (-7.0)

performance on all 4 datasets. On MNIST and Fashion-MNIST, Neural SFC model is trained for each class label. FFHQ has no labels, therefore all images are used together to learn a Neural SFC. On the TGIF dataset, Neural SFC model is trained on all gifs but generates a unique SFC for each gif, *i.e.*, all frames in a gif share the same Neural SFC.

The last column of Table 1 shows the average file size required to store an image in the test data and its relative improvement compared to the Raster scan. Specifically, in the TGIF section, the number represents the average size to save a single frame instead of the entire gif file. It is worth noting that this size does not include the order itself. One order can be shared by many images, thus the cost of it can be amortized. We observe that the Neural SFC results in smaller a sequence size on all 4 datasets, showing consistent improvement on the Raster scan or Hilbert Curve. Interestingly, as compared to the Raster scan, even though Hilbert Curve results in higher autocorrelation (see Figure 6), it is always worse in terms of LZW encoding length. This finding suggests that there is no simple relation between an order’s autocorrelation performance and LZW encoding length performance.

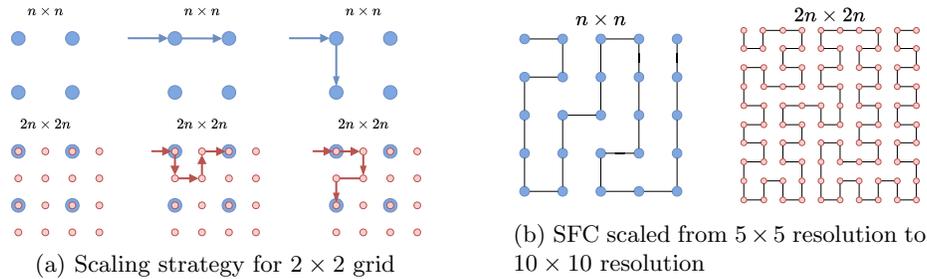


Fig. 7: **Scaling up SFC.** The top row (blue circles) shows a 2×2 crop of an image grid of resolution $n \times n$. The bottom row (red circles) shows how we can scale the SFC path from $n \times n$ grid to $2n \times 2n$ grid. For every incoming SFC to a pixel location, there are 3 possible ‘next’ pixels, straight ahead, left, or right. Column 2 and 3 consider the cases where the SFC goes ‘straight’ or ‘right’. The image on the right shows a toy example of scaling up SFC for a 5×5 image.

4.6 Scaling up SFC

Although in the current work, we only train Neural SFC models for 32×32 images to demonstrate their effectiveness, there are no theoretical restrictions of our approach for higher resolution images. It is trivial to either (1) train SFC for higher resolution itself or (2) scale an SFC computed for a low-resolution image to a high-resolution image, preserving the locality properties as shown in Fig. 7.

5 Conclusions and Future Work

We propose Neural SFC, the first data-driven approach to finding a context-based SFC for a set of images. We parameterize the SFCs as a set of weights over a graph defined using an image grid, and train a neural network to generate the weights. Neural SFC can be trained for any objective function defined for a set of images, even when not differentiable. We show the performance of Neural SFC on four real-world datasets on two different objective functions - (1) Pixel autocorrelations in the pixel sequences obtained from an image (2) Compressing a set of images or a short video such as a gif using LZW encoding.

While our work takes the first steps towards finding 1D sequences in 2D data, it opens up a number of directions for future research such as learning SFC for higher dimensional data such as 3D objects or in the case when the 2D space is in a latent space instead of pixels (VQ-VAE [29,34], dVAE [33]). Applying Neural SFCs to large video compression tasks [9,4] is also promising given their success on gifs. We will explore these exciting directions and more in future work.

Acknowledgements. This work was partially supported by the Amazon Research Award to AS.

References

1. Alexandrov, V., Alexeev, A., Gorsky, N.: A recursive algorithm for pattern recognition. *Proc. IEEE Intl. Conf. Pattern Recognition* pp. 431–433 (1982)
2. Ansari, A., Fineberg, A.: Image data compression and ordering using peano scan and lot. *IEEE Trans. on Consumer Electronics* **38**(3), 436–445 (1992)
3. Bader, M.: *Space-filling curves: an introduction with applications in scientific computing*, vol. 9. Springer Science & Business Media (2012)
4. Chen, H., He, B., Wang, H., Ren, Y., Lim, S.N., Shrivastava, A.: Nerv: Neural representations for videos. *Advances in Neural Information Processing Systems* **34**, 21557–21568 (2021)
5. Dafner, R., Cohen-Or, D., Matias, Y.: Context-based space filling curves. In: *Computer Graphics Forum*. vol. 19, pp. 209–218. Wiley Online Library (2000)
6. Dai, H., Khalil, E.B., Zhang, Y., Dilkina, B., Song, L.: Learning combinatorial optimization algorithms over graphs. *NIPS* (2017)
7. Deudon, M., Cournut, P., Lacoste, A., Adulyasak, Y., Rousseau, L.M.: Learning Heuristics for the TSP by Policy Gradient. In: *International conference on the integration of constraint programming, artificial intelligence, and operations research*. pp. 170–181. Springer (2018)
8. Drozdowski, M.: *Scheduling for parallel processing*, vol. 18. Springer (2009)
9. Ehrlich, M., Barker, J., Padmanabhan, N., Davis, L., Tao, A., Catanzaro, B., Shrivastava, A.: Leveraging bitstream metadata for fast and accurate video compression correction. *arXiv preprint arXiv:2202.00011* (2022)
10. Harary, F., Norman, R.Z.: Some properties of line digraphs. *Rendiconti del Circolo Matematico di Palermo* **9**(2), 161–168 (1960)
11. Hilbert, D.: Über die stetige abbildung einer linie auf ein flächenstück. In: *Dritter Band: Analysis: Grundlagen der Mathematik: Physik Verschiedenes*, pp. 1–2. Springer (1935)
12. Hopfield, J.J., Tank, D.W.: “neural” computation of decisions in optimization problems. *Biological cybernetics* **52**(3), 141–152 (1985)
13. Huffman, D.A.: A method for the construction of minimum-redundancy codes. *Proceedings of the IRE* **40**(9), 1098–1101 (1952)
14. Kamata, S., Eason, R.O., Kawaguchi, E.: An implementation of the hilbert scanning algorithm and its application to data compression. *IEICE TRANSACTIONS on Information and Systems* **76**(4), 420–428 (1993)
15. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4401–4410 (2019)
16. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *ICLR* (2016)
17. Kool, W., Van Hoof, H., Welling, M.: Attention, learn to solve routing problems! *ICLR* (2019)
18. Lawder, J.K.: Calculation of mappings between one and n-dimensional values using the hilbert space-filling curve. School of Computer Science and Information Systems, Birkbeck College, University of London, London Research Report BBKCS-00-01 August (2000)
19. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
20. Lee, J.H., Hsueh, Y.C.: Texture classification method using multiple space filling curves. *Pattern Recognition Letters* **15**(12), 1241–1244 (1994)

21. Lempel, A., Ziv, J.: Compression of two-dimensional data. *IEEE transactions on information theory* **32**(1), 2–8 (1986)
22. Li, Y., Song, Y., Cao, L., Tetreault, J., Goldberg, L., Jaimes, A., Luo, J.: Tgif: A new dataset and benchmark on animated gif description. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4641–4650 (2016)
23. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al.: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science* **326**(5950), 289–293 (2009)
24. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: *Proceedings of International Conference on Computer Vision (ICCV)* (December 2015)
25. Matias, Y., Shamir, A.: A video scrambling technique based on space filling curves. In: *Conference on the theory and application of cryptographic techniques*. pp. 398–417. Springer (1987)
26. Moon, B., Jagadish, H.V., Faloutsos, C., Saltz, J.H.: Analysis of the clustering properties of the hilbert space-filling curve. *IEEE Transactions on knowledge and data engineering* **13**(1), 124–141 (2001)
27. Moore, E.H.: On certain crinkly curves. *Transactions of the American Mathematical Society* **1**(1), 72–90 (1900)
28. Munroe, R.: xkcd: Map of the internet. <https://xkcd.com/195> (2006-12-11), accessed: 2021-11-16
29. Oord, A.v.d., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning. *NIPS* (2017)
30. Ouni, T., Lassoued, A., Abid, M.: Gradient-based space filling curves: Application to lossless image compression. In: *2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)*. pp. 437–442. IEEE (2011)
31. Peano, G.: Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen* **36**(1), 157–160 (1890)
32. Prim, R.C.: Shortest connection networks and some generalizations. *The Bell System Technical Journal* **36**(6), 1389–1401 (1957)
33. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I.: Zero-shot text-to-image generation. *ICML* (2021)
34. Razavi, A., Oord, A.v.d., Vinyals, O.: Generating diverse high-fidelity images with vq-vae-2. *NeurIPS* (2019)
35. Shannon, C.E.: A mathematical theory of communication. *The Bell system technical journal* **27**(3), 379–423 (1948)
36. Sierpinski, W.: Sur une nouvelle courbe continue qui remplit toute une aire plane. *Bull. Acad. Sci. Cracovie (Sci. math. et nat. Serie A)* pp. 462–478 (1912)
37. Thyagarajan, K., Chatterjee, S.: Fractal scanning for image compression. In: *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems & Computers*. pp. 467–468. IEEE Computer Society (1991)
38. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: *International Conference on Learning Representations* (2018), <https://openreview.net/forum?id=rJXMpikCZ>
39. Vinyals, O., Bengio, S., Kudlur, M.: Order matters: Sequence to sequence for sets. *ICLR* (2015)
40. Vinyals, O., Fortunato, M., Jaitly, N.: Pointer networks. *NIPS* (2015)

41. Welch, T.A.: Technique for high-performance data compression. *Computer* (1984)
42. Witten, I.H., Neal, R.M., Cleary, J.G.: Arithmetic coding for data compression. *Communications of the ACM* **30**(6), 520–540 (1987)
43. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017)
44. Yoo, A.B., Jette, M.A., Grondona, M.: Slurm: Simple linux utility for resource management. In: *Workshop on job scheduling strategies for parallel processing*. pp. 44–60. Springer (2003)
45. Zhou, L., Johnson, C.R., Weiskopf, D.: Data-driven space-filling curves. *IEEE transactions on visualization and computer graphics* **27**(2), 1591–1600 (2020)
46. Zhu, J., Hoorfar, A., Engheta, N.: Bandwidth, cross-polarization, and feed-point characteristics of matched hilbert antennas. *IEEE Antennas and Wireless Propagation Letters* **2**, 2–5 (2003)
47. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. *IEEE transactions on Information Theory* **24**(5), 530–536 (1978)