Eye of the Dragon : Exploring Discriminatively Minimalist Sketch-based Abstractions for Object Categories

Ravi Kiran Sarvadevabhatla
Video Analytics Lab
Supercomputer Education and Research Centre
Indian Institute of Science, Bangalore, India
ravikiran@ssl.serc.iisc.in

Venkatesh Babu R.
Video Analytics Lab
Supercomputer Education and Research Centre
Indian Institute of Science, Bangalore, India
venky@serc.iisc.in

ABSTRACT

As a form of visual representation, freehand line sketches are typically studied as an end product of the sketching process. However, from a recognition point of view, one can also study various orderings and properties of the primitive strokes that compose the sketch. Studying sketches in this manner has enabled us to create novel sparse yet discriminative sketch-based representations for object categories which we term *category-epitome*s. Concurrently, the epitome construction provides a natural measure for quantifying the sparseness underlying the original sketch, which we term epitome-score. We analyze *category-epitome*s and epitome-scores for hand-drawn sketches from a sketch dataset of 160 object categories commonly encountered in daily life. Our analysis provides a novel viewpoint for examining the complexity of representation for visual object categories.

Categories and Subject Descriptors

I.5.4 [**Applications**]: Computer Vision

General Terms

Experimentation

Keywords

freehand sketch; object category recognition; deep learning

1. INTRODUCTION

A master painter in ancient China was once painting an elaborate dragon. His son, carefully observing his father's craft, noticed that the eye of the dragon had not been painted. When he pointed it out, the painter replied: 画龙点睛 (paraphrased: "Adding the eye will make the dragon come alive and fly out of the painting").

A fascinating spectrum – from realistic depictions to sparsely drawn sketches – exists for hand-drawn art depicting objects[27, 3]. Across this spectrum, the level of detail in the artwork usually correlates with the confidence in recognizing the object. However, an interesting phenomenon is observed for freehand sketches : even with minimal stroke detail (see Figure 1), the underlying subject can be easily recognized by us. This suggests an inherent sparseness driving the human neuro-visual representation mechanism. Therefore, studying such sparsely detailed sketches can aid our understanding of the cognitive processes involved and spur the design of efficient representations and visual classifiers for objects.

Figure 1: A sketch belonging to category `cup`. In spite of minimal detail, we can recognize the sketch easily and correctly.

Typically, freehand line sketches are studied as an end product of the sketching process. Indeed, sketches have been utilized *in-toto* in the context of classification and content-based retrieval problems [11, 13, 22]. However, from a recognition point of view, we believe it is instructive to study the primitives (strokes) that compose the sketch, starting with the first hand-drawn stroke until the last stroke which finalizes the sketch. Our belief originates in a discovery we have made : for a given sketch of an object and an associated sketch classifier, there exists a minimal subset of strokes which ensures consistent and correct identification of the object category. We term this new sketch, constructed using the minimal stroke subset, as a *category-epitome*. Figure 2 shows examples of freehand line sketches and their corresponding sparse *category-epitome*s. In this paper, we explore the multi-faceted ways in which *category-epitome*s convey insights into the nature of visual object category representation. In the spirit of the dragon story mentioned at the beginning, we wish to examine sketches at the moment they come "alive".

MM '15 October 26-30, 2015, Brisbane, Australia

DOI: http://dx.doi.org/10.1145/2733373.2806230 ...$15.00.

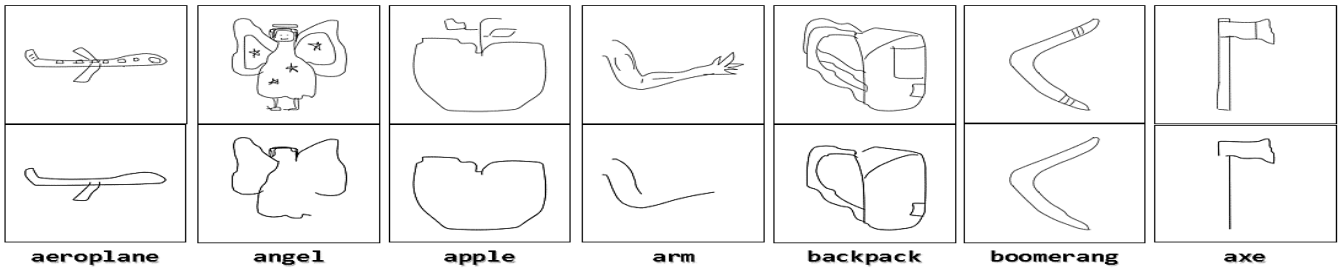

Figure 2: Original sketches (top row) and corresponding *category-epitome*s (bottom row) for various object categories.

We summarize our contributions below:

- Via *category-epitome*s, we introduce a novel representation for visual object categories. The freehand sketch of an object is already a sparse representation of the object category. *Category-epitome*s take this sparsification to the extreme while guaranteeing recognizability (Section 4). This guarantee provision is a unique feature which sets our method apart from other methods of sparse sketch generation and sketch simplification[19, 22, 24, 20].

- We show that *category-epitome*s can be constructed based on various philosophies of visual abstraction (Section 4.4). Our analysis of the resulting epitomes (Section 5.1) suggests that the longest sketch strokes best capture the "signature" sparsity of object categories in general.

- The measure we use to quantify the sparseness underlying a sketch is simple to understand and arises naturally out of its *category-epitome* construction (Section 4.3). Our analysis demonstrates that the distribution of epitome-scores can be used to quantify semantic level of detail in object categories (Section 5.2).

- To represent sketches, we propose novel features obtained from Alexnet, a popular deep-learning network [15]. Utilizing these features, we achieve state-of-the-art results for sketch classification on a challenging sketch dataset (Section 3.3).

2. RELATED WORK

The notion of a sparse, sketch-like representation for generic images was proposed in the seminal work of Marr [18] who coined the term *primal sketch* for the same. Primal sketch-like representations have been used as features for object detection [13], texture characterization [9] and for super resolution [25]. The idea of generating sketches from one or more source images, without any reference to the notion of primal sketch, has been explored by Qi et al. [22] and Marvaniya et al. [19]. In all these works, photographic images serve as the starting point. In contrast, our starting point is the sketch stroke data generated by human beings. This creates the possibility that the sparse neuro-visual representation is transferred to the sketch in the process of drawing by humans, at least in part. Though we do not survey the domain for reasons of space and focus, a significant body of work from Computer Graphics has also examined simplification of artistic renderings, including line art. The work of Ravish et al. [20] on abstraction of 3D-geometric models can be considered a representative example.

In addition to the artificial (i.e. not generated by human hand) nature of sketch generation, the works mentioned above do not attempt to quantify the sparseness of the resulting sketch nor do they examine the temporal nature of sketch composition. In contrast, recent work by Berger et al. [12] analyzes the temporal aspect of sketching in the context of mimicking artist style. However, their emphasis is on synthesizing abstract facial sketches rather than recognition. Moreover, their sketches are produced by professional artists. In contrast, our sketches have been generated by crowdsourcing and therefore, are more representative of object depictions by humans in general. The idea of identifying and utilizing a discriminative subset of strokes was employed by Karteek et al. [14] for classifying online handwritten data. The work of Eitz et al. [3] examines how humans tend to draw objects by analyzing a large number of sketches spread across commonly encountered object categories. We use their database of sketches in this paper.

The basic premise of sparse, sketch-like drawings being adequate for visual category recognition has also been acknowledged by the neuroscience community. As early as 1960, Gollin [7] used a series of drawings containing progressively more information as a measure of visual development and to examine long-term memory in amnesic patients. Dirk et al. [26] performed a functional MRI based study of participants viewing photographs and line drawings of natural scene categories. Their interesting conclusion was that the neural activity of the visual system in response to viewing line drawings is virtually the same as when coloured photographs are viewed, which indicates that line drawings capture the most important elements of a scene. Ghosh and Petkov [5] studied the robustness of contour based shape recognition algorithms to different kinds of contour incompleteness. While their work bears some resemblance to ours, the stroke segments they use are artificially created and do not correspond to natural breaks since temporal information is not captured. Moreover, as with some of the works we previously mentioned, their sketch-like representations are artificially generated, i.e. not generated by human hand.

3. BUILDING THE SKETCH CLASSIFIER

We first describe the sketch database used in our experiments. Subsequently, we describe sketch feature extraction (Sec. 3.2) and conclude the section with a discussion on the performance of our sketch classifier (Sec. 3.3). In the next section (Sec. 4), we describe how the sketch classifier is actually used to construct *category-epitome*s.

Figure 3: An ambiguously drawn sketch (left) with two plausible categorizations - `tyre` (middle) and `donut` (right). The figure is from the work of Schneider et al. [23].

3.1 The sketch database

The sketches in our study have been taken from the publicly available freehand line sketch database of Eitz et al. [3]. This database contains a curated set of $20,000$ hand-drawn sketches evenly distributed across 250 object categories. These sketches have been obtained by crowdsourcing across the general population. Therefore, they are a good starting point for analyzing the underpinnings of the sketching process by humans. Significantly, the temporal stroke information (i.e. the sequential order in which the strokes were drawn) for sketches is also available. As we will see subsequently, this plays a crucial role in obtaining *category-epitome*s involving temporal stroke ordering. A few examples from the sketch database can be seen in the top row of Figure 2.

While the database of Eitz et al. is quite comprehensive and general in coverage of object categories, it has its share of shortcomings, a number of which have been analyzed by Schneider et al. [23]. One of the major shortcomings is the presence of ambiguously drawn sketches whose identity is difficult to discern even for fellow human beings (See Figure 3. Is the sketch on the left depicted by the two concentric shapes a `tyre` or a `donut` ?). To address this situation, [23] employ a human-evaluation based technique and identify a subset containing 160 non-ambiguous object categories which can be utilized as a more reliable benchmark database for evaluating sketch recognition systems (Refer to Section 5 of [23] for details). For our experiments and analysis, we utilize the sketches from this curated set of 160 object categories. Furthermore, following [23], we uniformly consider 56 sketches from each category for the purpose of sketch classifier training and evaluation.

To increase the number of sketches per category for classification and improve performance[1], we perform data augmentation by applying geometric and morphological transformations to each sketch. Specifically, each sketch is initially subjected to image dilation ("thickening") using a 5×5 square structuring element. A number of transforms are applied to this thickened sketch – mirroring (across vertical axis), rotation ($\pm 5, \pm 15$ degrees), systematic combinations of horizontal and vertical shifts ($\pm 5, \pm 15$ pixels), central zoom ($\pm 3\%, \pm 7\%$ of image height). As a result, 30 new sketches are generated per original sketch. The data augmentation procedure results in $30 \times 56 = 1680$ sketches per category, for a total of $1680 \times 160 = 268,800$ sketches across 160 categories.

In the context of the sketches being processed by CNNs, the reason for sketch dilation must be pointed out. As each sketch gets processed by deeper layers of the CNN, fine details tend to get eliminated. To minimize the impact of detail loss, the sketches are subjected to thickening (dilation) which helps preserve detail better.

3.2 Feature Extraction and Classification

The sketch features are extracted using pre-trained Convolutional Neural Networks (CNNs). For our experiments, we utilize the pre-trained Alexnet CNN as described by Krizhevsky et al [15]. Our choice was motivated by the fact that features extracted from Alexnet have resulted in impressive performance across a variety of challenging computer vision problems [6, 8], surpassing the performance of hand-crafted features. The sketch features are obtained by tapping the output of layer conventionally denoted as fc_7 of the pre-trained Alexnet with the sketch as the input. Eventually, for an input sketch image, a 4096-dimensional feature vector is obtained.

The entire database of sketches is divided into training and test sets. Features from training set sketches are passed to a multi-class linear Support Vector Machine (SVM) classifier [4] whose parameter C has been empirically set to 0.5. The principled manner of choosing the SVM parameters would be to try different kernels and related parameter settings via cross-validation. Serendipitously, however, our choice of parameter C resulted in state-of-the-art sketch classifier performance (Section 3.3). Therefore, in order to focus on our objective of constructing *category-epitome*s and analyzing epitome properties, we deferred the task of kernel and parameter selection.

3.3 Evaluation of sketch classifier

For comparable evaluation, we utilize the methodology and test set described in [23]. We summarize this procedure next. To begin with, the curated dataset contains 160 sketch categories, each containing 56 sketches. Since some of the categories contain more than 56 sketches, we randomly select a 56-sized subset from these categories. The number of sketches utilized for training is progressively increased, starting from 8 sketches per category in steps of 8 up to 48 of the 56 sketches. The rest of the sketches are utilized for testing[1]. The entire data is randomly shuffled thrice and for each shuffle, it is split according to one of the training and testing splits mentioned above. For each shuffle, precision is calculated over test data. The resulting precision values are then averaged. This procedure of shuffling thrice and computing average precision is repeated for each of the 8 train/test splits considered. For the SVM classifier, parameter C is empirically set to 0.5.

The results can be seen in Figure 4. Our Alexnet feature-based recognition system consistently outperforms the hitherto best results in [23] – the improvement in average precision ranges from 3% to 11%. It must be conceded that the performance of the method in [23] is based on unaugmented data. However, they employ Fisher vector feature representation whose dimensions[2] are larger than our 4096-dimensional feature vectors by a factor of about 10. This would result in prohibitively large memory requirements and training times. In contrast, our feature vectors are faster to

[1]Note that for training, data augmented variants (Section 3.1) of each sketch are used while testing is done only on the original sketch subjected to dilation and not on the data augmented variants. As an example, when 32 of the original sketches are used, the actual number of training sketches is $32 \times 30 \times 160 = 153,600$ while the number of test sketches is $24 \times 160 = 3840$.

[2]The Fisher vector feature dimension is estimated based on the description of feature extraction process in [23].

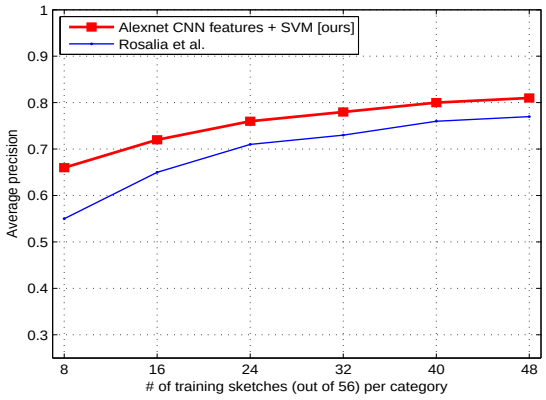

Figure 4: Comparing the performance of our sketch classifier with that of Schneider et al. [23]. The x-axis shows the number of sketches (subset size) used for training.

process. Therefore, they are also a good candidate for related applications such as sketch-based image retrieval [11].

A classifier with good performance, as in our case (Figure 4), translates to a larger percentage of sketches from the test set being classified correctly. This, in turn, translates to a larger set of *category-epitome*s for analysis since the epitomes can only be constructed for correctly classified sketches, as we will see shortly. In the next section, we shall see how this sketch classifier is utilized in the construction of *category-epitome*s.

4. OBTAINING THE *CATEGORY-EPITOME*

As the first step in determining the *category-epitome*, for each correctly classified test sketch, we construct a cumulative sketch sequence from its strokes (Section 4.1). Each of the sketches in this sequence is labeled using the sketch classifier. The resulting sequence of labels is analyzed to determine the *category-epitome* of the sketch. Our procedure for obtaining the *category-epitome* also provides a natural measure, termed epitome-score, for quantifying the sparseness of the original, full sketch (Section 4.3).

The temporal sequence in which the strokes are added is, by default, the basis for construction of sketch image sequences of cumulative strokes. However, apart from temporal stroke order, other plausible stroke orderings exist. These stroke orderings optionally utilize the length of strokes as well in the construction of *category-epitome*s. Towards the end of this section (Section 4.4), we describe how these additional stroke ordering based sketch image sequences are created.

4.1 Constructing Cumulative Stroke Sequences

As the first step in determining the *category-epitome*, we construct sequences of cumulative strokes derived from correctly classified test sketches. Suppose the sequence of strokes in the temporal order they were drawn in a sketch is given by $S_t = \{s_1, s_2 \ldots s_N\}$ where N is the total number of strokes in the sketch. To construct the corresponding cumulative stroke sequence, we begin with a blank canvas. Strokes from the given sketch are successively added to the blank canvas in the temporal order. As each stroke is added, intermediate canvases $S_1, S_2, \ldots, S_N$ are created. Specifically, the intermediate canvases are given by $S_1 = \{s_1\}, S_2 = \{s_1, s_2\}, \ldots, S_N = \{s_1, s_2, \ldots, s_N\}$. At the end of this process, we obtain the cumulative stroke sequence $CSS_T = \{S_1, \ldots, S_N\}$ for the temporal stroke order. Figure 5 illustrates the creation of cumulative stroke sequence for a sketch from `airplane` category.

4.2 Constructing the *category-eptiome*

Having generated the cumulative stroke sequence for a sketch as described above, the *category-epitome* can be constructed. To begin with, each intermediate canvas of the cumulative stroke sequence is classified to obtain a binary labeling – the label is 1 if the sketch category is correctly identified by the sketch classifier and 0 otherwise. Thus, we obtain a binary label sequence $\mathcal{L} = \{l_1, l_2, \ldots, l_N\}$ corresponding to each canvas of the cumulative stroke sequence CSS_T (see Figure 5).

Note that the final canvas S_N corresponds to the original test sketch since all the strokes have been added to the canvas at that point. Therefore, the final classification label l_N must be 1 since we are working with correctly classified test sketches. Now, consider the product sequence $\mathcal{P} = \{P_1, P_2 \ldots . P_N\}$ formed by cumulative multiplication of labels $l_i \in \mathcal{L}, i = 1, 2 \ldots N$:

$$P_i = \prod_{j=i}^{N} l_j \tag{1}$$

Then, the *category-epitome* corresponds to canvas S_e of the cumulative stroke sequence such that

$$e = \min_{1 \le i \le N} \{i | P_i = 1\} \tag{2}$$

Informally, the *category-epitome* S_e is the sequentially earliest, correctly classified canvas whose successors are classified correctly as well. Using the example in Figure 5, the classification label sequence is $\mathcal{L} = \{0, 1, 0, 1, 1, 1, 1, 1, 1\}$. From Equation (1), the product sequence is computed as $\mathcal{P} = \{0, 0, 0, 1, 1, 1, 1, 1, 1\}$. From Equation (2), we obtain $e = 4$. In other words, canvas S_4 (outlined by a cyan rectangle in Figure 5) corresponds to the *category-epitome*: the sequentially earliest, correctly classified canvas whose successors $S_5 \ldots S_9$ are classified correctly as well. Figure 2 shows sketches from various categories and their corresponding *category-epitome*s.

4.3 Epitome-score : Quantifying the *category-epitome*

Our procedure for obtaining the *category-epitome*, described above, also provides a natural method for quantifying the "epitome"-ness of the original, full sketch. Using e obtained from Equation (2), we define the epitome-score $\mathcal{E}$ of a sketch as :

$$\mathcal{E} = \begin{cases} \frac{e}{N}, e \neq 1 \\ 0, \quad e = 1 \end{cases} \tag{3}$$

where N is the total number of strokes in the sketch. $e = 1$ corresponds to the situation where merely drawing the first stroke conveys the epitome-ness of the sketch. Therefore, for consistency across sketches, we define the corresponding

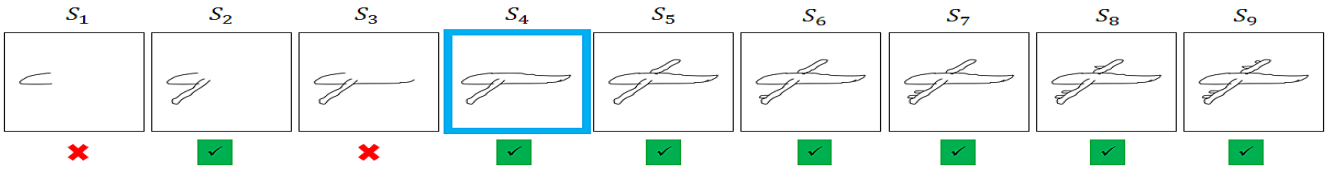

Figure 5: Constructing the TEMPORAL *category-epitome* for an `airplane` sketch: The sketch has 9 strokes. $S_1 - S_9$ are the cumulative stroke sequence canvases. A red cross mark indicates a misclassification(0) while a green tick mark indicates correct classification(1). Canvas S_4 outlined by a cyan rectangle is the *category-epitome*. Note that even though canvas S_2 is classified correctly, we consider *category-epitome* as canvas S_4, the sequentially earliest, correctly classified canvas whose successors are classified correctly as well.

epitome-score $\mathcal{E}$ to be 0. Our definition of epitome-score $\mathcal{E}$ essentially conveys the sparseness underlying the sketch – the smaller its value, the more sparser the sketch is likely to be. Epitome-scores very close to 1, on the other hand, indicate that very few strokes in the original sketch are perceptually irrelevant. Fortunately, very high epitome-scores are not the norm, as we will see – sparsity is pervasive across categories. Referring once again to Figure 5, the epitome-score for the `airplane` sketch is computed as $\mathcal{E} = \frac{4}{9} = 0.44$ (see Equation 3).

4.4 Additional Stroke Sequence Orderings

The temporal sequence in which the strokes are added is, by default, the basis for construction of cumulative stroke sequence. However, apart from temporal stroke order, we propose other plausible stroke orderings. For completeness, we include the previously introduced temporal stroke ordering as well.

1. TEMPORAL : Strokes are accumulated in the order they were drawn temporally (see top row of Figure 6). Using the notation for sketch strokes defined in Section 4.1, the intermediate canvases are given by $S_1 = \{s_1\}, S_2 = \{s_1, s_2\}, \ldots, S_N = \{s_1, s_2, \ldots, s_N\}$. The resulting sketch sequence $CSS_T = \{S_1, S_2 \ldots, S_N\}$ reflects the hypothesis that the natural order in which people draw sketch strokes best captures the epitome-ness of the sketch.

2. LENGTH : Strokes are accumulated in the order of non-increasing length, i.e. longer strokes are accumulated before adding the shorter ones(second row of Figure 6). Suppose the sorted-by-non-increasing-length version of the original temporal sequence S_t is $S_l = \{s_{l_1}, s_{l_2} \ldots s_{l_N}\}$. Then, the intermediate canvases are $S_1 = \{s_{l_1}\}, S_2 = \{s_{l_1}, s_{l_2}\}, \ldots S_N = \{s_{l_1}, s_{l_2}, \ldots, s_{l_N}\}$. The resulting sequence $CSS_L = \{S_1, S_2 \ldots S_N\}$ reflects the hypothesis that longest strokes best capture the epitome-ness of the sketch.

3. ALTERNATE : Strokes are accumulated by alternately adding strokes of non-increasing length and strokes considered in reverse temporal order (see third row of Figure 6). In this case, the intermediate canvases are $S_1 = \{s_{l_1}\}, S_2 = \{s_{l_1}, s_N\}, S_3 = \{s_{l_1}, s_N, s_{l_2}, \ldots\}$ which in turn generate $CSS_A = \{S_1, S_2 \ldots, S_N\}$. Note that in constructing the sequence, strokes are removed from S_t(temporal sequence) and S_l(non-increasing length sequence) upon being added to CSS_A. This sequence reflects the hypothesis that the longest strokes *and* decorative elements (finishing touches) typically added towards the end of the drawing best capture the epitome-ness of the sketch.

The stroke sequence orderings essentially capture various philosophies of visual abstraction as manifested in *category-epitome*s. Once the cumulative sketch sequence (CSS_T or CSS_L or CSS_A) of a correctly classified full sketch is obtained, the procedure for determining the *category-epitome* (Section 4.2) and epitome-score (Section 4.3) is essentially the same.

The full set of *category-epitome*s for all three stroke sequence orderings can be viewed at `http://val.serc.iisc.ernet.in/eotd/epitome_images`.

5. ANALYSIS

For our experiments, we utilize the classifier obtained by training with 24 (out of 56) sketches from each category. The ensuing accuracy of the classifier is 76% (see Figure 4). While it is certainly possible to consider many other training and test splits of data, we chose a split which provides a classifier with acceptable performance while maximizing the potential number of *category-epitome*s obtained.

5.1 Epitome-scores across categories for various sketch orderings

We begin our analysis by computing the median of epitome-scores for correctly classified sketches on a category-by-category basis. The median epitome scores are computed for all the three sketch orderings described in Section 4.4 viz. TEMPORAL, LENGTH and ALTERNATE. The median epitome scores along with corresponding error bars are presented in Figure 8 for correctly classified test sketches across the 160 object categories.

The first heartening aspect is that most of the median-epitome scores are small (i.e. not close to 1). As Figure 7 shows, the percentage of categories which have median epitome-scores below 0.5 ranges between 57% and 87% for various sketch stroke orderings. These trends suggest the viability of *category-epitome*s. More significantly, they suggest an inherent sparseness for visual representation of object categories because the smaller the epitome-score, the sparser the *category-epitome* sketch typically.

If we examine the scores closely, the epitome-scores for some of the categories (E.g. `apple` and `nose` of TEMPORAL sketch order in Figure 8, topmost row, leftmost column) are 0. For these categories, the sketches are typically drawn such that a dominant stroke or two essentially captures the epitome-ness of the category (See also Equation 3).

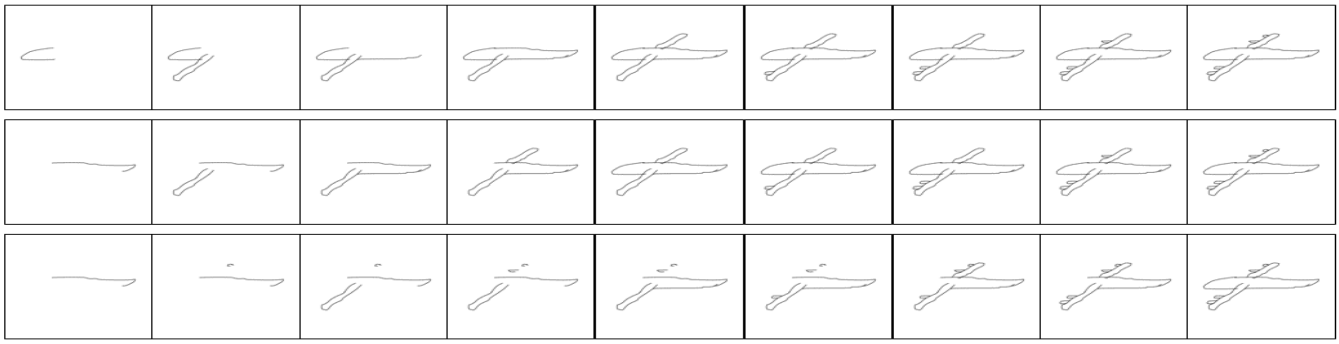

Figure 6: A sketch from `airplane` category and its cumulative stroke sequence canvases for the stroke orderings TEMPORAL (top row), LENGTH (middle row) and ALTERNATE (bottom row).

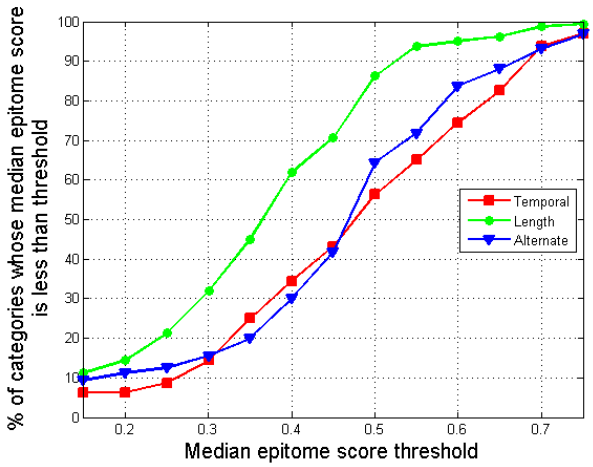

Figure 7: Most of the median epitome-scores are small: The figure shows the % of total categories whose median epitome-score is below a certain threshold. The plots are depicted for the three sketch orderings.

The varying lengths of error bars in Figure 8 indicates the variability in depiction across categories. For example, the sketches of the category `sun` (topmost row, middle column in Figure 8) are drawn with a fairly consistent appearance by humans. This consistency influences the sketch classifier and can cause it to produce *category-epitome*s which exhibit a minor amount of variations, thus resulting in a compact distribution of epitome-scores (shorter error bars). In contrast, the variety in sketches of a category such as `flying bird` (same plot as that of `sun`) is reflected in the corresponding *category-epitome*s and by extension, in the longer error bars of its epitome-score.

If we compare the median epitome scores across sketch orderings (Figures 7 and 8), it is easy to see that the LENGTH sketch ordering (adding strokes in decreasing order of length) produces the sparsest epitomes. The result is very interesting since it runs contrary to intuition – we expect people to somehow draw the most discriminative strokes first (i.e. in TEMPORAL stroke ordering). Thus, temporal information does not seem to play any role in obtaining the sparsest epitomes. This throws the door open for analysis of additional freehand sketch datasets where temporal information is absent [21, 16]. The results also demonstrate that the next viable ordering (ALTERNATE) is not good enough. An interleaving of longest strokes with the decorative strokes (typically added temporally towards the end of the sketching process) does result in sparser epitomes compared to temporal order, but an even sparser yet recognizable sketch can be obtained if the strokes are added in order of decreasing length.

5.2 Epitome-score as a proxy for semantic level of detail

A different perspective can be gained by examining categories for a given value of epitome-score. For each category, we count the number of test sketches whose epitome-score is less than a threshold and normalize by the number of test sketches in the category. To facilitate analysis and avoid visual clutter, we select 10 prototypical categories for display. The plot that results by varying the threshold can be viewed for various sketch sequencing orders in Figure 9. The 10 categories are chosen as follows : The AUC (area-under-the-curve) is computed for each category's plot. The AUCs are sorted across categories. Finally, the categories corresponding to the top 4, bottom 4 and the middle 2 AUC values are chosen for representation.

The epitome-score can be considered as a proxy for semantic level-of-detail. Viewed in this light, the plots from Figure 9 suggest the varying epitome-score budgets across categories. Some categories require a considerable level of detail before their epitomal avatars are revealed. These are typically categories whose plots are towards the lower right in the figures (E.g. `monkey` and `angel` in Figure 9). On the other hand, categories with plots towards the upper left corner (E.g. `apple` and `nose` in Figure 9) have relatively less stringent demands on level of detail.

6. DISCUSSION

In this paper, we have presented a preliminary exploration of *category-epitome*s as a sparse yet discriminative representation of visual object categories. Our analysis of *category-epitome*s and their epitome-scores provides a novel viewpoint for studying the complexity of representation for object categories. Our method of constructing the *category-epitome* has a unique feature which sets it apart from other

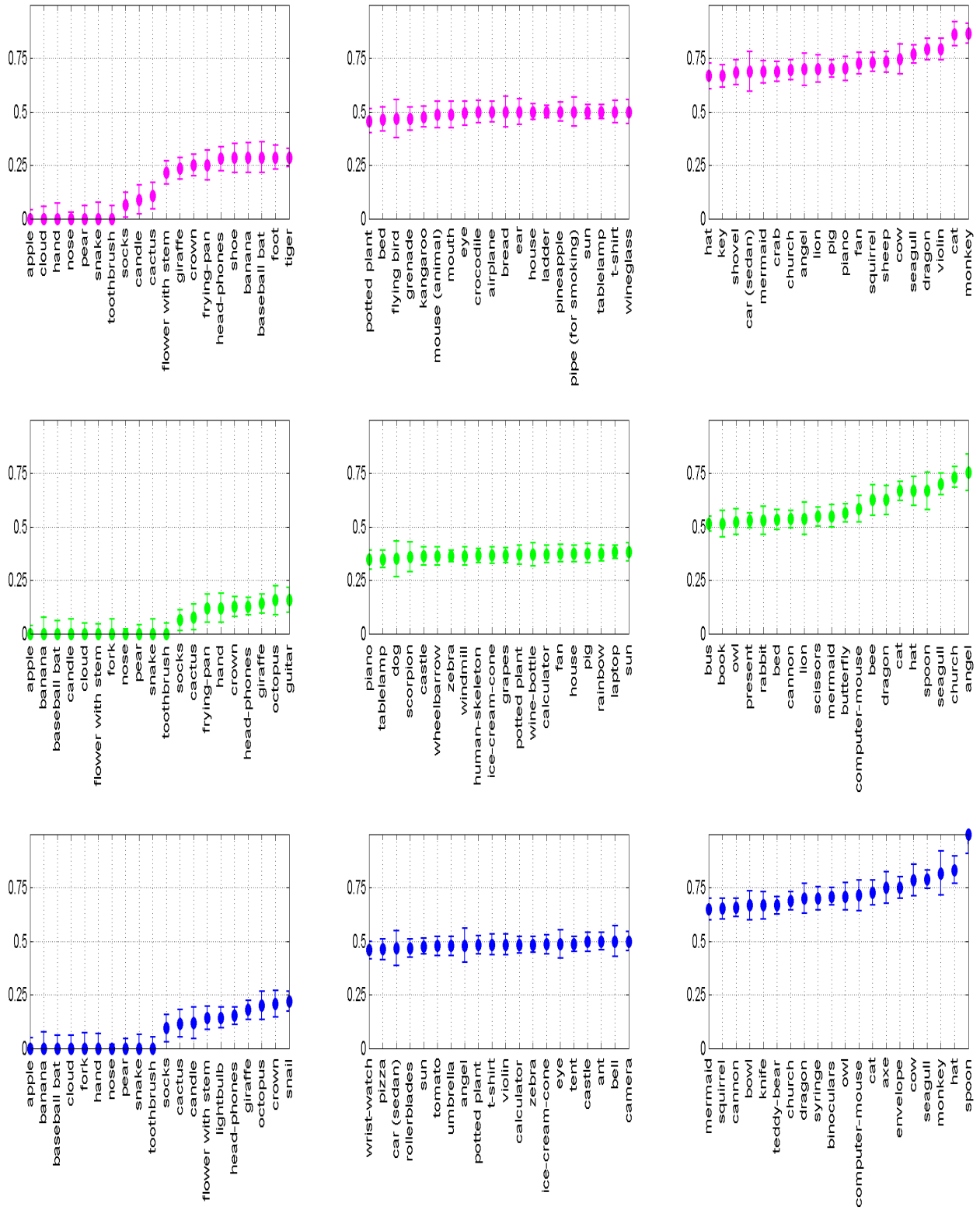

Figure 8: Sorted median epitome-scores (y-axis) and corresponding error bars for 160 object categories(x-axis). For clarity, only the first, middle and last 20 categories of the sorted order are shown. The standard errors are clamped to $[0, 1]$ – the range of epitome-scores. The median scores are shown as filled circles. The plots above depict the trends for the three different philosophies of abstraction – TEMPORAL (topmost plot, magenta), LENGTH (middle plot, green) and ALTERNATE (bottom plot,blue). Plots for the full set of 160 categories can be viewed at http://val.serc.iisc.ernet.in/eotd/.

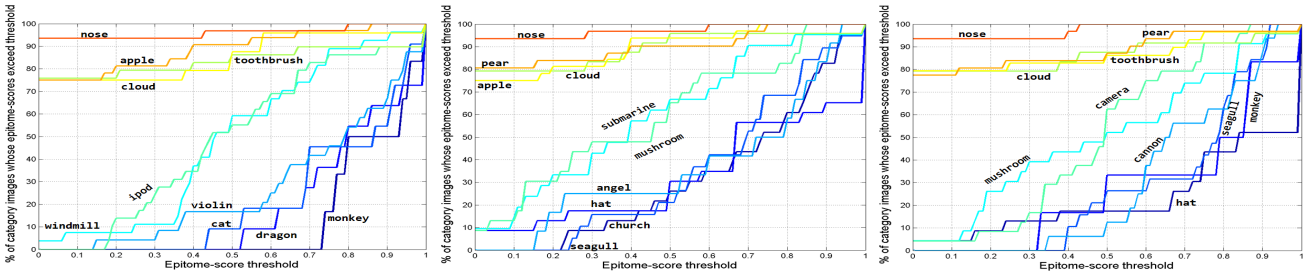

Figure 9: Demonstrating the effect of budgeting epitome-score on category recognition rate using data from 10 selected categories. The values of epitome-score thresholds are shown on x-axis. The % of images in each category which exceed a particular value of the epitome-score threshold are shown on y-axis. The 10 category names are shown adjacent to their respective plots. The plots depict the trends for the three different philosophies of abstraction – TEMPORAL (left), LENGTH (middle) and ALTERNATE (right) .

methods of sparse sketch generation and sketch simplification [22, 19, 24, 20]: It guarantees that the most fundamental primitives (strokes) which enable category recognition (discrimination) are retained. Since the construction of *category-epitome*s renders them incomplete yet recognizable sketches, they can also be used to benchmark the robustness of sketch features and sketch classification pipelines. In addition, epitome-score of the *category-epitome*s can serve as a quantitative measure for evaluation of Non-Photorealistic Rendering methods, a need sorely felt by the researchers active in the area [17, 10, 2].

A practical application domain for *category-epitome*s is the field of neuro healthcare. Partially drawn sketches of objects are used to analyze the extent of visual agnosia: a neurological disorder in which object recognition is impaired in humans [7]. The epitome-score provides an additional measure for medical practitioners to monitor and characterize the disorder.

Comparison with the work of Schneider et al. [23]: Suppose a sketch has n strokes. Schneider et al. [23] determine a stroke's contribution (positive, negative, no effect) on object recognizability by removing the stroke and retaining the remaining $(n-1)$ strokes. However, to truly determine stroke importance, other (smaller than $(n-1)$ strokes) subsets must also be considered. For example, the petal strokes of a flower sketch, when removed individually, may be grouped as negative contributors. However, a smaller subset of strokes which contains both petals and other strokes can result in correct classification, as our results suggest. Also, the analysis by [23] is qualitative. We additionally provide a novel quantitative measure (epitome-score) for sketches. Finally, our method also results in sparse yet recognizable representation for objects.

7. FUTURE WORK

A number of other directions exist for future work. One such would be to compare the performance across various Convolutional Neural Networks and analyze the effect of tapping features from different layers of such networks. More generally, we hope that our work will spur interest in analysis of sketches utilizing cutting-edge tools of the deep learning methodology.

By the nature of their generation, sketches have an inherent sequential structure (of strokes). The classifier we use does not exploit this information. Therefore, another direction of work would be to utilize classifiers similar in spirit to Hidden Markov Models (HMM) which can exploit such information for improved sketch recognition. On a related note, it would be worthwhile to explore if classifiers whose performance is relatively superior produce sparser epitomes.

In our work, we have analyzed *category-epitome*s arising out of three different methods of visual abstraction (stroke ordering). The methods are by no means exhaustive. Other methods, such as those based on stroke curvature and randomized stroke data occlusion [5] are potential candidates for constructing *category-epitome*s as well. However, we defer the related study.

In this paper, the results and analysis rely on a machine classifier trained on sketches. A natural extension would be to analyze *category-epitome*s obtained when humans are asked to identify sketches given only a subset of strokes from the original sketch. Comparing the resulting epitomes and epitome-scores would delineate the differences between humans and machines in terms of the cognitive load imposed by partial sketch recognition. In fact, the epitomes and intermediate predictions generated by humans and the machine classifier could be used to conduct a Turing test to quantify the aforementioned difference. To this end, we have deployed a web app where users are randomly shown sketches in different stroke sequence orderings and asked to recognize the sketch. At the time of writing, the web app has been successfully used by about 500 people and has enabled us to collect recognition data for nearly 9000 sketches.

Given their minimalist appearance, *category-epitome*s can also be considered "stylized" yet recognizable depictions of objects. To evaluate the style factor as perceived by humans, we are currently developing an app which tracks user preferences for various philosophies of sketch abstraction (stroke orderings). The resulting statistics would reflect the degree of human agreement on style-quotient of machine-generated abstract art.

8. ACKNOWLEDGMENTS

We wish to acknowledge the anonymous reviewers for their feedback. In addition, we thank the members of Video Analytics Lab, IISc for their feedback during various stages of the draft preparation. Ravi Kiran Sarvadevabhatla would like to acknowledge Victor Ng-Thow-Hing for relating anecdotes which served to inspire the paper's title and the opening story.

9. REFERENCES

[1] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *BMVC*, 2014.

[2] F. Cole, K. Sanik, D. DeCarlo, A. Finkelstein, T. Funkhouser, S. Rusinkiewicz, and M. Singh. How well do line drawings depict shape? In *SIGGRAPH*, 2009.

[3] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *SIGGRAPH*, 2012.

[4] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, 2008.

[5] A. Ghosh and N. Petkov. A cognitive evaluation procedure for contour based shape descriptors. *Int. J. Hybrid Intell. Syst.*, 2(4):237–252, Dec. 2005.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[7] E. S. Gollin. Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*, 11:289–298, 1960.

[8] Y. Gong, L. Wang, R. Guo, and S. Lazebnik. Multi-scale orderless pooling of deep convolutional activation features. In *ECCV*. 2014.

[9] R. Haralick, L. T. Watson, and T. J. Laffey. The topographic primal sketch. *The International Journal of Robotics Research*, 2(1):50–72, March 1983.

[10] A. Hertzmann. Non-photorealistic rendering and the science of art. In *NPAR*, 2010.

[11] R. Hu and J. Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *CVIU*, 117(7):790–806, July 2013.

[12] I.Berger, A. Shamir, M. Mahler, E. Carter, and J. Hodgins. Style and abstraction in portrait sketching. *SIGGRAPH*, 2013.

[13] I.Kokkinos, P. Maragos, and A. Yuille. Bottom-up and top-down object detection using primal sketch features and graphical models. In *CVPR*, 2006.

[14] A. Karteek, P. S. Lahari, and C. V. Jawahar. Discriminant substrokes for online handwriting

[14] A. Karteek, P. S. Lahari, and C. V. Jawahar. Discriminant substrokes for online handwriting recognition. In *ICDAR*, 2005.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[16] L. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, 2000.

[17] H. Li, D. Mould, and J. Davies. Structure and aesthetics in non-photorealistic images. In *Proceedings of Graphics Interface*, 2013.

[18] D. Marr. Chapter 2: Representing the image. In *Vision*, pages 54–79. The MIT Press, 2010.

[19] S. Marvaniya, S. Bhattacharjee, V. Manickavasagam, and A. Mittal. Drawing an automatic sketch of deformable objects using only a few images. In *ECCV. Workshops and Demonstrations*, 2012.

[20] R. Mehra, Q. Zhou, J. Long, A. Sheffer, A. Gooch, and N. J. Mitra. Abstraction of man-made shapes. In *ACM SIGGRAPH Asia*, 2009.

[21] S. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-20). Technical report, 1996.

[22] Y. Qi, J. Guo, Y. Li, H. Zhang, T. Xiang, and Y. Song. Sketching by perceptual grouping. In *ICIP*, 2013.

[23] R. G. Schneider and T. Tuytelaars. Sketch classification and classification-driven analysis using fisher vectors. *SIGGRAPH Asia*, 2014.

[24] A. Shesh and B. Chen. Efficient and dynamic simplification of line drawings. *Comput. Graph. Forum*, 27(2):537–545, 2008.

[25] J. Sun, N.Zheng, H.Tao, and H. Y. Shum. Image hallucination with primal sketch priors. In *CVPR*, 2003.

[26] D. B. Walther, B. Chai, E. Caddigan, D. M. Beck, and L. Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *PNAS*, 108(23):9661–9666, 2011.

[27] W. Zhang, X. Wang, and X. Tang. Lighting and pose robust face sketch synthesis. In *ECCV*, 2010.