

Laconic Image Classification: Human vs. Machine Performance

Javier Carrasco
DCC, Universidad de Chile & IMFD
Santiago, Chile
jcarrasc@dcc.uchile.cl

Aidan Hogan
DCC, Universidad de Chile & IMFD
Santiago, Chile
ahogan@dcc.uchile.cl

Jorge Pérez
DCC, Universidad de Chile & IMFD
Santiago, Chile
jperez@dcc.uchile.cl

ABSTRACT

We propose *laconic classification* as a novel way to understand and compare the performance of diverse image classifiers. The goal in this setting is to minimise the amount of information (aka. *entropy*) required in individual test images to maintain correct classification. Given a classifier and a test image, we compute an approximate minimal-entropy positive image for which the classifier provides a correct classification, becoming incorrect upon any further reduction. The notion of entropy offers a unifying metric that allows to combine and compare the effects of various types of reductions (e.g., crop, colour reduction, resolution reduction) on classification performance, in turn generalising similar methods explored in previous works. Proposing two complementary frameworks for computing the minimal-entropy positive images of both human and machine classifiers, in experiments over the ILSVRC test-set, we find that machine classifiers are more sensitive entropy-wise to reduced resolution (versus cropping or reduced colour for machines, as well as reduced resolution for humans), supporting recent results suggesting a texture bias in the ILSVRC-trained models used. We also find, in the evaluated setting, that humans classify the minimal-entropy positive images of machine models with higher precision than machines classify those of humans.

CCS CONCEPTS

- **Information systems** → **Multimedia information systems**;
- **Computing methodologies** → **Neural networks**.

KEYWORDS

image classification; laconic classification; deep neural networks

ACM Reference Format:

Javier Carrasco, Aidan Hogan, and Jorge Pérez. 2020. Laconic Image Classification: Human vs. Machine Performance. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, October 19–23, 2020, Virtual Event, Ireland. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3340531.3411984>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '20, October 19–23, 2020, Galway, Ireland

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6859-9/20/10...\$15.00
<https://doi.org/10.1145/3340531.3411984>

1 INTRODUCTION

Deep neural networks now surpass human-level performance on various specific tasks relating to visual recognition. In a widely-used yardstick for human-level performance, Russakovsky et al. [22] estimated that an expert human trained on the task can achieve a top-5 classification error rate of 5.1% on a dataset of 1,500 ILSVRC images and 1,000 target classes. Shortly after, He et al. [8] surpassed human-level performance on the same task achieving 4.9% top-5 error with PReLU-net. Later works would further reduce this error rate, including ResNet50 (3.6%) [9], Trimps-Soushen (3.0%) [23], SeNetResNet50 (2.3%) [15], etc., with contemporary state-of-the-art models more than halving estimated human error for this task.

Though such results represent landmark advances for machine vision, by focusing on classification errors over high-quality images alone, they do not reveal the full story of relative machine performance for image classification. Works on adversarial examples [1, 20, 26], for instance, establish that human and machine perception diverges greatly for specifically constructed images. Other works have presented bespoke experiments comparing human and machine performance beyond classification errors, presenting evidence for a lack of robustness in the presence of noisy [2, 3, 22] or incomplete information [12, 19, 24, 27, 28, 31], a sensitivity to spatial [4, 5, 10, 30] or colour [13, 14] transformations, a lack of generalisation [7], a bias towards texture [6], etc., in the machine classifiers studied. By transforming test images prior to classification, these works provide insights into the differing types of information that humans and machines rely on for image classification.

These latter recent works suggest the need for an information-theoretic framework that generalises such issues: a framework within which the performance of classifiers – be they human, machine or other – can be compared and understood, allowing to quantify, in a more fine-grained manner, the *type of information* in the input on which a given classifier depends. While previous works address individual or multiple types of information reduction on input images in isolation, a more general framework should allow to combine and compare different types of reduction on test inputs.

Overview. In this paper, we propose an intuitive information theoretic framework for understanding classification results based on the principle of computing and analysing *minimal entropy positive inputs*: inputs with minimal information with respect to yielding correct classification results. The notion of entropy generalises and allows for comparing the relative effects of different reductions on inputs – and their combinations – on classifier performance. Such reductions may include, for example, downsampling, quantisation and slicing. The goal in this framework thus shifts from precise classification to *laconic classification*: providing a classifier that minimises the entropy of input(s) required for correct classification. Using a continuous notion of entropy rather than a discrete notion of

correct/incorrect introduces a novel challenge beyond minimising classification error (for which state-of-the-art approaches already achieve near-perfect results on ILSVRC). Existing datasets – such as ILSVRC – can be directly used for evaluating classifiers under this new goal. Models performing well for laconic classification should likewise perform well in practical settings involving incomplete or noisy information capture (low light, distant objects, etc.).

Though the framework we propose can in principle be applied to any classifier for any classification task, herein we first instantiate the framework on the aforementioned problem of image classification. We consider three general operations for reducing the entropy of test images: crop (slicing), resolution reduction (downsampling) and colour reduction (quantisation). We then propose two methods for finding the minimal entropy positive images under these reductions for two different types of classifier.

Given a pre-trained machine classifier – where classification can be separated from learning – an input test image, and a set of reduction operations, we apply the given reduction functions to the input image to find (under certain assumptions) the lowest entropy image that the model classifies correctly such that applying any further reduction of entropy leads to incorrect classification. We apply the aforementioned framework to find the minimal entropy images from a sample ILSVRC test-set for state-of-the-art deep neural-network (DNN) models (GOOGLENET, SQUEEZE NET, RESNET50 and SENETRESNET50), with respect to the three aforementioned reduction operations and their combination.

We compare these results with human classifiers. Applying our framework for humans is not trivial since learning cannot be separated from classification (we cannot start with the full input test-image and reduce it since the human will remember the image) and automated search is not possible. We thus design a method *reversing* the optimization goal: starting from a void image, the human evaluator may add information incrementally until they believe that they can classify the image. We apply this framework with more than 500 human users through an online interface.

Finally, with the goal of ascertaining how characteristic are the minimal entropy positive images of a classifier (i.e. how different is the type of information that different classifiers need for successful classification) we cross-classify the minimal-entropy positive images among DNNs and humans: given classifier A and B , we present A 's minimal-entropy positive images to B and vice versa, computing the traditional measure of classification precision.

Use-cases. We envisage a number of use-cases for laconic classification. Firstly, as explored in this paper, it can serve as a performance *metric* for comparing the quality of classification provided by different models. Secondly, it can serve as a way to *explain* the classification results of different models by producing images that illustrate the minimal information required for correct classification. Thirdly, the entropy required for correct classification can become an *objective* towards which models can be optimised, which may improve their performance in low-information settings. Fourthly, the framework may be useful for *compressing* input data used by classifiers, throwing away information that is not required for correct classification at the source, potentially saving bandwidth or other resources. Herein we primarily focus on the first two use-cases in

order to initially understand which classifiers require more/less information, and what kinds of information they depend on.

Contributions. Our main contributions are as follows:

- (1) We propose a framework – based on minimal entropy positive inputs – within which classifiers can be evaluated, optimised and compared. The framework can be adapted to different types of classifiers and classification tasks.
- (2) Given a machine classifier, a set of labelled inputs, and a set of entropy reductions, we propose a top-down method – starting from a full-quality input – and a bottom-up method – starting from a void input – for computing the minimal entropy positive inputs of a given classifier.
- (3) We present experiments in the setting of image classification over the ILSVRC dataset, where we compare the amount of information required by humans and machines for correct classification with respect to three forms of entropy reduction, and their combination.
- (4) We provide a set of minimal entropy positive images for both machine and human classifiers. We believe that the human-generated images, in particular, can serve as a relevant benchmark for comparing machine models with human-level performance in terms of (laconic) image classification.

Summary of results. As a summary of our observations:

- Our experiments show that minimal entropy images are considerably smaller than original images for DNNs; for example, we find that with only 2–6% of the information content of the original test-images (on average) the best performing machine model can still produce a correct classification.
- The minimal-entropy positive images for humans tend to be considerably smaller than their machine equivalents in the case of colour and resolution reductions (54–59% of the information required by the best performing machine model). On the other hand, in the case of crop, humans tend to require more information than machine models (143% of the information required by the best performing machine model).
- The precision of human classifiers on machine minimal-entropy positive images for machines was considerably better (0.74 precision in the worst case) than the corresponding results for cross-classification of human-generated minimal-entropy positive images by machine models (0.29 precision in the worst case for the best model).

These observations support the following results:

- State-of-the-art machine models can correctly classify images from smaller cropped regions than those from which humans can perform correct classification.
- Humans perform better for image classification in low colour and low resolution settings than these machine models.
- Amongst these machine models, good performance for laconic classification correlates with good performance for classification tasks on full-quality images.

The first result supports previous observations of a lack of robustness with regards to incomplete information for image classification relative to humans in state-of-the-art machine models [3], and a bias towards texture in ILSVRC-trained models [6].

Known limitations. The limitations of the setting in which our results have been developed are as follows:

- Our results are computed for a subset of ILSVRC images with a reduced set of classes. Other datasets may be considered.
- We consider three forms of entropy reduction for images and their combination. Other reductions may be considered.
- We measure entropy based on the Portable Network Graphics (PNG) image format. Other measures may be considered.
- We currently experiment with machine models that have been pre-trained and optimised for classification of full-quality images. Better results may be obtained from machine models with specialised training for laconic classification.
- The method for computing minimal images in humans and machines is (necessarily) different. Human error may lead to latent classification ability being under-estimated.

The first three limitations could be straightforwardly addressed in future work using the proposed methods. The fourth limitation raises an interesting open question: how should models be trained for laconic classification in low-information settings? The fifth limitation appears more fundamental, stemming from the different ways in which machines and human process information.

Paper structure. The paper is structured as follows. Section 2 presents related works on the generation of minimal images for the classification task, on the robustness of machine classifiers for images, and on the comparison of human and machine classification performance. Section 3 defines the notion of a minimal entropy positive input, instantiates this notion for three types of entropy reduction, and proposes two distinct methods to approximate the minimal entropy positive inputs of classifiers. Section 4 discusses the experimental setting in terms of the models used, the datasets and classes used, etc. Section 5 presents the results of our experiments, comparing the entropy reduction for the minimal entropy positive images of different classifiers, and measuring the precision for cross-classification of these images. Section 6 concludes with a summary of the main results, and directions for future work.

2 RELATED WORKS

We provide an overview of works relating to the computation of minimal images for successful human and machine classification; and the robustness (or lack thereof) of machine models to certain types of noise, information reduction, and/or input variations.

Minimal Images. Previous works have proposed notions of minimal images with respect to the image classification task. Ullman et al. [27] introduced the notion of a “minimal image” as the smallest region of an image that is still recognisable by a human. They show that a small change in these minimal images can have a drastic effect on recognition. Moreover, Ullman et al. [27] show that DNNs are unable to accurately recognise human minimal images. To compute the minimal images, they started from the complete image and then iteratively cropped it showing the new cropped image to a different human subject every time. The process stopped when the accuracy of recognizing every crop of the current region dropped below a threshold. Given the difficulty of computing minimal regions recognisable by humans, their experiments were limited to 10 images [27]. Srivastava et al. [24] extended the previous work

by showing that the sharp drop in accuracy for minimal images can also be observed in DNNs. They defined the notion of “fragile recognition image” as a region of an image for which a small change in size produces a considerable change in the accuracy of recognition by DNNs. They showed that fragile recognition images are abundant and can occur at different sizes. Zhang et al. [31] compare human and machine performance for image classification over segments of an image based on object boundaries (rather than rectangular regions). Both human and machine classifiers are used to identify key segments for classification, where, interestingly, humans are found to be better at classifying images using segments selected by machine models versus those selected by humans.

Robustness. Another line of related works study the robustness of DNN image classification with respect to distortions to the input image [2–7, 10, 11, 13, 14, 30]. These works consider the performance of image classification subject to distortions including pixel noise, defocusing, blur, over-saturation, rotations, inversions, spatial transformations, etc. Works by Geirhos et al. [7] and Dodge and Karam [3] include comparison with human performance in such settings, showing that the performance of DNNs is much lower than human performance on distorted images. Geirhos et al. [7] further show that the performance of DNNs can be improved by specifically training on images that include the various distortions, but that the resulting models struggle to generalise, performing poorly when presented with new types of noise. In a different line of research, works by Hosseini and Poovendran [13] and Xiao et al. [30] explore adversarial examples, which involve making minimal changes (sometimes even imperceptible to humans) to an image such that the classifier no longer provides the expected result.

Novelty. Works by Ullman et al. [27] and Srivastava et al. [24] define minimal images in terms of minimal contiguous regions of images (under crop) that still yield image classification. We generalise this idea by proposing minimal images in terms of the entropy of the image, i.e., the amount of information contained in the image. Under this entropy-based framework, image cropping then becomes one form of entropy reduction that we explore, alongside reductions in resolution, colour, and combinations thereof; our framework also generalises to other forms of reduction. On the other hand, while the works that look at classification performance under various forms of distortion are (like us) concerned more generally with the robustness of classifiers [2–7, 10, 11, 13, 14, 30], individual distortions are considered orthogonal, where robustness is explored along different dimensions in isolation. Our work can be seen as exploring robustness along one dimension – entropy – which generalises the two lines of work discussed here, and can be used as a unifying measure under which various types of reductions and distortions can be considered. However, entropy does not capture – nor does it intend to capture – all possible distortions. For example, two transformations considered by Geirhos et al. [7] – image rotation and colour inversion – may not alter the amount of information; other transformations that they consider – such as Eidolon and additive noise – may even increase the image file size. In contrast we currently only consider transformations that reduce the information content of the original input image file.

Table 1: Alternatives considered for entropy measure, indicating compression rate from best (1) to worst (5) according to Larkin [18], the level of support in standard libraries, and limitations that prevent use for our framework

Rate	Measure	Support	Limitations
1	WebP	Medium	—
2	PNG	High	—
3	delentropy	—	Single channel / 8 bit
4	GIF	High	256 Colour Palette
5	JPEG-LS	Low	—

3 APPROXIMATING MINIMAL-ENTROPY POSITIVE IMAGES

We now discuss the framework we use for evaluating the laconic classification of images. We first discuss the entropy measure and the reductions applied to images in this work. We then discuss the computation of approximate minimal-entropy positive images (MEPIs) for DNNs and for humans.

3.1 Entropy Measure

Our goal is to find the minimal-entropy positive images (MEPIs) for various classifiers based on various forms of entropy reduction. Although there do exist entropy measures proposed for images that are loosely inspired by analogous Shannon-like probabilistic measures for strings [18, 29], such measures of entropy are complicated by the conditional and joint entropy within image regions and across channels, and thus do not support important features, such as colour. Alternatively, we can consider a Kolmogorov complexity [17] for images, whereby, given a (Turing complete) descriptive language \mathcal{L} , the entropy of an image I will be measured in terms of the length of the shortest bit string $S_I \in \mathcal{L}$ (an algorithm in the language) that produces I when evaluated; the effect of the language chosen on the complexity measure has a constant bound. Unfortunately, no general-purpose encoder exists to compute S_I from I .¹ This leaves us with the practical alternative of estimating entropy in terms of a lossless compression scheme $(E, D) \in \mathcal{L} \times \mathcal{L}$, with a fixed encoder E and a fixed decoder D such that $D(E(I)) = I$ for all $I \in \mathcal{I}$, the set of images supported. The algorithmic entropy of the image with respect to (E, D) is then defined simply as $|E(I)|$: the size (in bits) of the losslessly compressed image.

Along these lines, in Table 1 we present the alternatives considered for estimating the entropy of images. With the exception of delentropy [18], which is a probabilistic entropy measure, the alternatives refer to lossless image compression algorithms. We rule out delentropy and GIF due to limitations regarding the dimensionality of images supported. Of the alternatives, we identify JPEG-LS, PNG and WebP as viable, where we choose PNG due to its widespread availability. Importantly, since we will use entropy as a relative rather than absolute measure, our framework should not be sensitive to the particular choice of entropy measure.

¹If a general-purpose encoder E were to exist – such that $E(I) = S_I$ – it would have a fixed number of bits, and there would then exist a fixed value for n such that we could include E in a program with n bits that enumerates inputs to E and outputs the first input with a Kolmogorov complexity greater than n : a paradox.

3.2 Entropy Reduction

We now define the entropy reductions considered. For simplicity we will assume that reductions are defined over matrices of non-negative integers, though the framework generalises straightforwardly to reals and other domains. Given an $m \times n$ matrix \mathbf{A} , we denote by a_{ij} ($1 \leq i \leq m$, $1 \leq j \leq n$) the element in the i^{th} row and j^{th} column of \mathbf{A} . We consider the following reductions:

Definition 3.1 (Quantisation). $\mathbf{A} \downarrow_{Q(\kappa)}$ is defined as the nearest-value $m \times n$ quantised matrix of \mathbf{A} with factor $0 \leq \kappa \leq \frac{\max(\mathbf{A})-1}{\max(\mathbf{A})}$, such that $a'_{ij} := \text{round}(\kappa \cdot a_{ij})$ for all a'_{ij} in $\mathbf{A} \downarrow_{Q(\kappa)}$.

Definition 3.2 (Downsampling). $\mathbf{A} \downarrow_{D(\sigma)}$ is defined as the $r \times s$ ($r \leq m$, $s \leq n$) downsampled matrix of \mathbf{A} with scaling factor σ such that $0 < \sigma < 1$, $\lfloor \sigma m \rfloor = r$, $\lfloor \sigma n \rfloor = s$, and $r < m$ or $s < n$.

Definition 3.3 (Slice). $\mathbf{A} \downarrow_{S(\alpha, \beta, \gamma, \delta)}$ is defined as the (contiguous) $p \times q$ submatrix such that $\mathbf{A} \downarrow_{S(\alpha, \beta, \gamma, \delta)} = (\mathbf{A}_{ij})_{\alpha < i \leq m - \beta; \gamma < j \leq n - \delta}$ where $\alpha, \beta, \gamma, \delta$ are non-negative integers ($\alpha + \beta < m$, $\gamma + \delta < n$, $p = m - \alpha - \beta$, $q = n - \gamma - \delta$, $\alpha + \beta + \gamma + \delta > 0$); in other words, the first α rows, the last β rows, the first γ columns and the last δ columns are removed from \mathbf{A} .

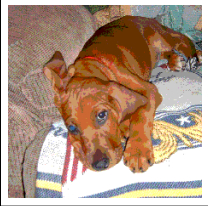
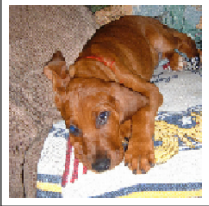
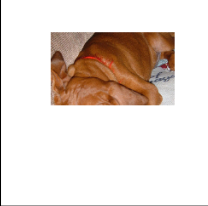
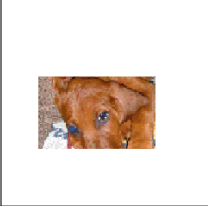
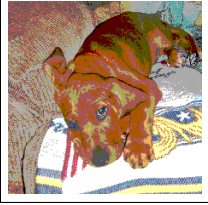
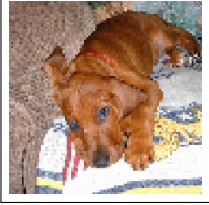
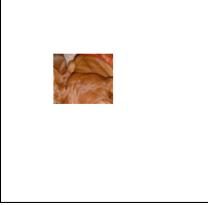
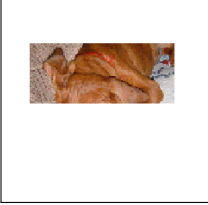
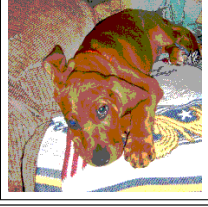
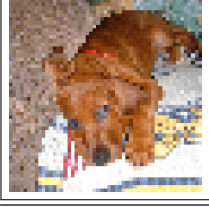
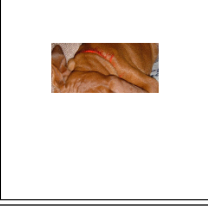
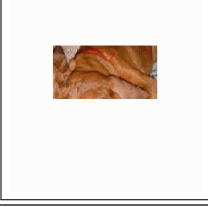
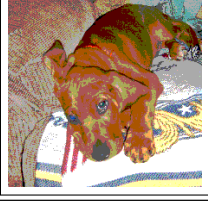

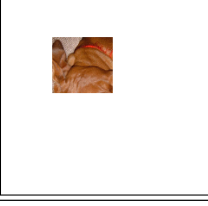
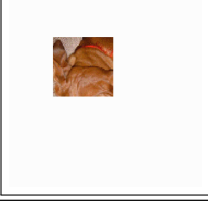
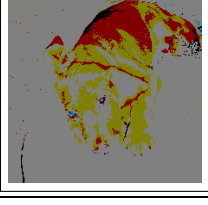

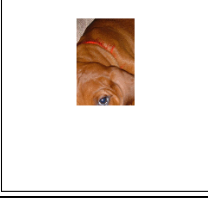
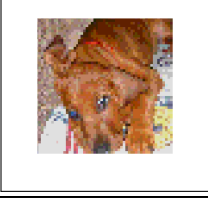
These reductions will be applied in atomic steps for which we assume parameters $\varepsilon_1, \varepsilon_2, \varepsilon_3$. We use $\mathbf{A} \downarrow_Q$ to denote $\mathbf{A} \downarrow_{Q(\varepsilon_1)}$, $\mathbf{A} \downarrow_D$ to denote $\mathbf{A} \downarrow_{D(\varepsilon_2)}$, and $\mathbf{A} \downarrow_S$ to denote $\mathbf{A} \downarrow_{S(\varepsilon_3, \varepsilon_3, \varepsilon_3, \varepsilon_3)}$. We also use $\mathbf{A} \downarrow_{S \nabla}$ to denote $\mathbf{A} \downarrow_{S(\varepsilon_3, 0, 0, 0)}$, $\mathbf{A} \downarrow_{S \Delta}$ to denote $\mathbf{A} \downarrow_{S(0, \varepsilon_3, 0, 0)}$, $\mathbf{A} \downarrow_{S \triangleright}$ to denote $\mathbf{A} \downarrow_{S(0, 0, \varepsilon_3, 0)}$ and $\mathbf{A} \downarrow_{S \triangleleft}$ to denote $\mathbf{A} \downarrow_{S(0, 0, 0, \varepsilon_3)}$, removing rows/columns from the top, bottom, left and right, respectively.

These three forms of entropy reduction, as applied to matrices, are general: lowering the precision of values in the matrix, computing a smaller matrix with new values encoding information from the full matrix, and slicing a matrix into a smaller sub-matrix with its original values. Likewise one might consider combining these reductions in arbitrary ways to form new reductions.

In the case of images, these reductions correspond, respectively, to colour reduction (parameter: κ), resolution reduction (parameter: σ), and crop (parameters: $\alpha, \beta, \gamma, \delta$). We also consider their combination, giving six parameters ($\alpha, \beta, \gamma, \delta, \kappa, \sigma$) by which to reduce entropy. We adapt these reductions to multi-channel images in the natural way. In the case of crop, we apply the reduction to all channels separately; however, based on initial experiments with DNNs, rather than remove the rows and columns of the image’s channels, we rather replace them with a constant neutral value, which allowed further entropy reduction in positive images by avoiding distortions once images are internally rescaled (i.e., images with narrow crops being “stretched out”). In the case of colour reduction, the nearest quantisation values are computed in the multi-channel case based on Euclidean distance; we further normalise the output values to fill the colour space after the quantisation, choosing equidistant points. In the case of downsampling, the reduction is applied to each channel and maintains the same aspect ratio.

In Table 2 we present examples of the aforementioned reductions on an image of a dog from the ILSVRC dataset [22]. In fact, these images correspond to the minimal entropy positive images for five classifiers, as we will discuss in the sub-sections that follow.

Table 2: The minimal-entropy positive images (MEPIs) computed from an example ILSVRC image of a dog for four DNN-based machine classifiers and humans considering colour, resolution and crop reductions, as well as their combination

Model	Colour	Resolution	Crop	Combined
SQUEEzENET [16]:				
GOOGLENET [25]:				
RESNET50 [9]:				
SENETRESNET50 [15]:				
HUMAN:				

3.3 Minimal-Entropy Positive Inputs

We now define a minimal-entropy positive input, where we assume a particular entropy measure H , denoting the entropy of a matrix A as $H(A)$. We also assume a classification task with a set of labels L and a ground-truth labelling λ such that $\lambda(A) \in L$. Let \mathcal{R} be a set of *reduction steps* (e.g., $\mathcal{R} = \{Q, D, S\triangleright, S\triangleleft, S\triangleright, S\triangleleft\}$). We say that the *reduction edge* $A' \xrightarrow{R} A''$ holds if and only if $A' \downarrow_R = A''$ and $H(A') > H(A'')$. The *reduction graph* $G_{A, \mathcal{R}}$ of A and \mathcal{R} is then the set of all reduction edges $A'' \xrightarrow{R} A'''$ that hold such that $R \in \mathcal{R}$ and either $A'' = A$, or there exists an edge $A' \xrightarrow{R'} A''$ in the reduction graph. Noting the requirement that $H(A') < H(A'')$, the reduction

graph is then a directed, acyclic, edge-labelled graph. Let $\mathcal{R}^*(A)$ denote the set of matrices A_n recursively reachable from A in $G_{A, \mathcal{R}}$ through a directed path of the form $A \xrightarrow{R_1} A_1 \xrightarrow{R_2} \dots \xrightarrow{R_n} A_n$ ($n \geq 1$); we include A in $\mathcal{R}^*(A)$. We begin by defining an initial notion of a minimal-entropy positive input that we later refine.

Definition 3.4 (Naive minimal-entropy positive input). Given a matrix A , a set of entropy reductions steps \mathcal{R} , a classifier C , and a ground truth labelling λ , we define the naive minimal-entropy positive input of A with respect to \mathcal{R}, C, λ as the matrix:

$$\arg \min_{A' \in \{A'' \in \mathcal{R}^*(A) \mid C(A'') = \lambda(A)\}} H(A').$$

This definition may give undesirable results in cases where a classifier may simply “guess” a correct label from a void input; for example, an image classifier making predictions based on the individual (nondescript) pixels of the input image might “guess” the correct class for a pixel, which would become a naive minimal-entropy positive image. To avoid such cases, we add a continuity condition: that there exists a continuous path of reduction edges from the input matrix through (only) correctly classified matrices. More formally let $\mathcal{R}_{\lambda,C}^*(\mathbf{A})$ denote the set of matrices \mathbf{A}_n recursively reachable from \mathbf{A} in $G_{\mathbf{A},\mathcal{R}}$ through a directed path of the form $\mathbf{A} \xrightarrow{R_1} \mathbf{A}_1 \xrightarrow{R_2} \dots \xrightarrow{R_n} \mathbf{A}_n$ ($n \geq 1$) such that $C(\mathbf{A}_i) = \lambda(\mathbf{A})$ for $1 \leq i \leq n$. As a special case, if $C(\mathbf{A}) = \lambda(\mathbf{A})$ then \mathbf{A} is included in $\mathcal{R}_{\lambda,C}^*(\mathbf{A})$; otherwise $\mathcal{R}_{\lambda,C}^*(\mathbf{A})$ is defined as the empty set. We can then define a *monotonic minimal-entropy positive input*.

Definition 3.5 (Monotonic minimal-entropy positive input). Given a matrix \mathbf{A} , a set of entropy reductions steps \mathcal{R} , a classifier C , and a ground-truth labelling λ , we define the monotonic minimal-entropy positive input of \mathbf{A} with respect to \mathcal{R}, C, λ as the matrix:

$$\arg \min_{\mathbf{A}' \in \mathcal{R}_{\lambda,C}^*(\mathbf{A})} H(\mathbf{A}')$$

For image classification, we refer to monotonic minimal-entropy positive inputs as minimal entropy positive images (MEPIs).

3.4 Approximating MEPIs for DNNs

In the case of DNNs, for a given image, classifier, ground-truth label and applicable reductions, to search for the MEPI, we incrementally applying atomic reduction steps to the image while the prediction of the classifier remains correct, backtracking in the case of an incorrect prediction. While straightforward for reductions with a single parameter, in the case of crop or combined reductions, we have multiple parameters over which we must search; for example, in the case of crop, for an $m \times n$ image, the number of images to check in the worst-case is potentially $\binom{m}{2} \binom{n}{2}$, i.e., more than 274 billion contiguous sub-images for a 1024×1024 input image (assuming step sizes of one pixel). Approximation is thus sought.

First, our assumption of continuity helps us to prune the search space: if we reach a set of parameter values for which classification is correct but for which any further atomic reduction is incorrect, we know we can rule out all further reductions from that point. We may still, however, encounter an infeasible search space when considering multiple parameters; to improve performance, we choose to apply a greedy search algorithm: given that the function we wish to minimise is not differentiable but has a fixed number of inputs, we apply Powell’s method [21] to find a local minimum.

In Figure 2, we provide the MEPIs computed for an example input image of a dog using the method described, considering four DNN-based machine classifiers (described later) with respect to the three aforementioned reductions, and their combination.

3.5 Approximating MEPIs for Humans

Humans operate differently to DNNs and thus require specialised methods to approximate their MEPIs. First, humans cannot separate classification from learning; hence we cannot show the full input image and apply reductions as the human will (of course) remember

the full input image. Second, humans require (much) more time per classification and can suffer from fatigue given complex tasks.

Given \mathbf{A} , a set of reduction steps \mathcal{R} , a set of labels L and a ground truth labelling λ , we compute a bottom-up approximation of MEPIs for humans. We start with a *void matrix* in the reduction graph $G_{\mathbf{A},\mathcal{R}}$ – a matrix for which further reduction is not possible, and thus without outgoing edges – where the search is applied in the reverse direction of the edges, increasing the entropy towards the original matrix. At each step, for the current matrix \mathbf{A}'' , the classifier must choose to either (1) select an available reduction R , such that $\mathbf{A}' \xrightarrow{R} \mathbf{A}''$, where \mathbf{A}' becomes the current matrix; or (2) pick a label for the current matrix \mathbf{A}'' from L . When the original matrix \mathbf{A} is reached, the classifier may pick a label or pass. The search ends once a label is picked. If the label is correct ($\lambda(\mathbf{A})$), the current matrix is returned as the minimal-entropy positive input; otherwise the search is inconclusive and the image is passed.

When navigating the reduction graph in the reverse direction, steps may be non-deterministic; for example, given a void matrix with a single value v , navigating backwards over $S\Delta$ may yield a 2×1 matrix for each value of v with some value below it in \mathbf{A} . Hence we restrict the search to make it deterministic. In practice we start with the void image representing the central pixel of the original image. To mitigate fatigue, larger steps are defined such that no more than 20 are required to return to the input image along a given dimension. In the case of multiple reductions, we offer the option to undo one step of any reduction or all reductions at once. These simplifications make it feasible to approximate MEPIs for humans, but it is important to note that with the limitations of coarser steps and always starting at the central pixel, the MEPIs approximated using this method are more likely to overestimate the entropy required for correct classification.

In Figure 2, we provide example MEPIs generated by a human for an input image of a dog using the bottom-up method described.

4 EXPERIMENTAL SETTING

We now discuss the experimental setting, describing the images used, and the DNN-based machine classifiers selected.

Data: We use images from the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [22]. As discussed by Russakovsky et al. [22], the 1,000 detailed classes of ILSVRC (e.g., COUCAL, SEALYHAM TERRIER) require expert training for humans to perform adequately at classification, not only to visually distinguish the objects, but also to retrieve their label from the 1,000 options. Thus, along similar lines to experiments conducted by Geirhos et al. [7] and Zhang et al. [31], we frame the task for the following twenty high-level classes: BEAR, BIRD, CAT, DOG, FISH, FLOWER, FOX, FRUIT, FUNGUS, HIPPOPOTAMUS, INSECT, LION, MONKEY, REPTILE, SHARK, SPIDER, TIGER, VEGETABLE, VEHICLE, WOLF. We select these classes as they should be generally recognisable to humans without prior training; furthermore, we select mostly plants and animals to provide a more challenging classification task, with visually similar classes, such as LION/TIGER, FRUIT/VEGETABLE, DOG/WOLF, INSECT/SPIDER, etc., providing non-trivial cases to visually distinguish. The results of DNN models are then mapped hierarchically to these higher-level classes (e.g., COUCAL \mapsto BIRD,

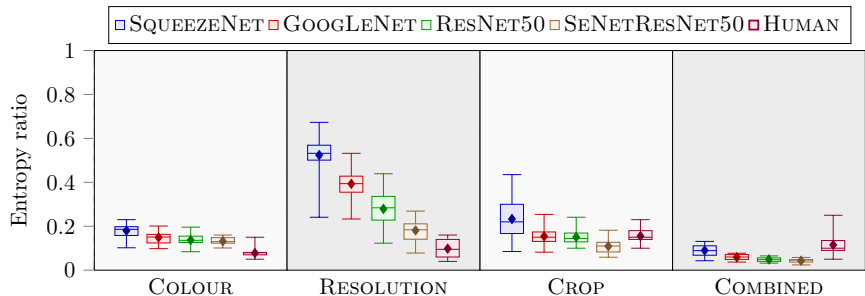


Figure 1: Box-plots of mean entropy ratio for DNN and human MEPIs across classes

SEALYHAM TERRIER \mapsto DOG). We sample 15 images for each of the 20 classes from the ILSVRC2012 test set for experiments.

Machine classifiers: We select four DNN classifiers trained on the ILSVRC2012 training set for the purposes of our experiments. These four classifiers – presented here in order of their performance for images classification in the traditional setting over the ILSVRC dataset – are SQUEEZE NET [16], GOOGLE NET [25], RES NET 50 [9] and SE NET RES NET 50 [15]. As discussed in the introduction, RES NET 50 and SE NET RES NET 50 have been found to surpass human-level performance for top-5 classification error on 1,500 ILSVRC images and 1,000 target classes. We refer to Table 2 for an initial impression of the MEPIs generated by the four models for an example image.

5 EXPERIMENTAL RESULTS

With our experimental results, we address two main questions: (1) How does the entropy required by DNN and human classifiers compare? (2) How do the classifiers perform in terms of precision for each others’ MEPIs? We address these two questions in the following subsections, along with other underlying questions relating to how different image reductions affect individual models, how the goals of laconic and accurate classification correlation, etc.

5.1 Entropy Ratio in MEPIs

For the $20 \times 15 = 300$ test images, we first compute the top-down MEPIs for the four DNN classifiers using the method described in Section 3.4. For humans, following the bottom-up method described in Section 3.5, we provide a web interface that – starting with a void image – allows a human user in each step to either enhance the image by the given dimensions, or select the class for the currently displayed image. In order to achieve many responses, the interface was shared on a university forum as well as on social media. The options and instructions were presented in Spanish, corresponding to the native language of the country in which the university is based. In total, 423 user sessions were logged; of the 1,722 responses obtained (average 4.07 per session), 1,340 (77.8%) were correct and thus yielded valid human-generated MEPIs across the 20 classes.

In Figure 1 we present the *entropy ratio* for four different settings across five different classifiers; entropy ratio is defined here as the ratio of the (PNG-encoded) size of the original input image versus the size of the extracted MEPI. We take the average entropy ratio for the images of each of the twenty classes. Figure 1 then presents the box-plots – displaying the 1st (min), 25th (lower quartile), 50th

(median), 75th (upper quartile), and 100th (max) percentiles with the mean marked as a diamond – for the mean ratio across the different classes; for example, in the case of SQUEEZE NET, considering only the COLOUR experiment, the best class had a mean entropy ratio of 0.10 (bottom whisker), the worst class had a mean entropy ratio of 0.23 (top whisker), the median class gave 0.19 (line inside the box), the lower and upper quartiles gave 0.16 and 0.20 (box edges), and the mean ratio for all classes was 0.18 (the diamond). We present the DNN models in order of their reported performance for top-5 classification error on the ILSVRC dataset, with SQUEEZE NET having the highest such error and SE NET RES NET 50 having the lowest.

Machine results. From Figure 1, we can draw some high-level observations about the DNN classifiers. First, for the DNN classifiers, the entropy ratio is lowest for the COMBINED experiment (as expected), which offers more avenues by which to reduce the entropy while maintaining a correct classification. Second, the results for DNN classifiers follow the same trend as for performance based on classification error; this suggests that there is a correlation between the goals of laconic classification and precise classification. Third, we see that DNNs are most sensitive to reductions in resolution, which supports the hypothesis that DNNs trained on ILSVRC images are biased towards texture [6] (with RESOLUTION being the parameter that most affects the ability to distinguish texture).

Human vs. machine results. With respect to human classifiers, we note that they are less sensitive to reductions in COLOUR (see Table 2 for an example) and much less sensitive to reductions in RESOLUTION than all DNNs, but more sensitive to reductions in CROP than some state-of-the-art DNNs. We also see an unusual result, whereby the reduction ratios for COMBINED are not lower than those for (e.g.,) COLOUR; this suggests that the users may have struggled with the interface for the CROP and COMBINED experiments, where multiple options were provided to enhance the image (versus COLOUR and RESOLUTION, which only permitted enhancements along one dimension). As such, the results for CROP and COMBINED in human classifiers leave an ambiguity: are the relatively poor results of humans due in this case to greater sensitivity to such (multi-dimensional) reductions, or because of difficulty using the more complex interface in these cases? We require further experiments to address this ambiguity.

	SN	GN	RN	SRN	Hum
SN	1.000	0.470	0.333	0.373	0.136
GN	0.943	1.000	0.647	0.647	0.255
RN	0.950	0.920	1.000	0.773	0.432
SRN	0.943	0.893	0.790	1.000	0.424
Hum	0.923	0.911	0.891	0.914	1.000

(a) COLOUR

	SN	GN	RN	SRN	Hum
SN	1.000	0.313	0.173	0.093	0.039
GN	0.897	1.000	0.283	0.120	0.030
RN	0.900	0.843	1.000	0.347	0.152
SRN	0.943	0.943	0.830	1.000	0.285
Hum	0.929	0.907	0.918	0.864	1.000

(b) RESOLUTION

	SN	GN	RN	SRN	Hum
SN	1.000	0.043	0.077	0.017	0.112
GN	0.600	1.000	0.220	0.037	0.221
RN	0.550	0.267	1.000	0.023	0.226
SRN	0.820	0.600	0.690	1.000	0.424
Hum	0.867	0.816	0.852	0.760	1.000

(c) CROP

	SN	GN	RN	SRN	Hum
SN	1.000	0.023	0.017	0.010	0.068
GN	0.457	1.000	0.107	0.020	0.092
RN	0.580	0.267	1.000	0.117	0.232
SRN	0.747	0.457	0.390	1.000	0.348
Hum	0.847	0.809	0.833	0.740	1.000

(d) COMBINED

Figure 2: Heatmaps of precision on MEPIs for the four types of entropy reduction; rows indicate the classifier, columns indicate the MEPIs: SN = SQUEEZE_{NET}; GN = GOOGLE_{NET}; RN = Res_{NET}50; SRN = Se_{NET}Res_{NET}50; Hum = HUMAN

5.2 Cross-classification precision for MEPIs of different classifiers

We turn to the precision of (top-1) classification across the MEPIs of models, again considering COLOUR, RESOLUTION, CROP and COMBINED. To gather results for human classification, we created a

second, simpler, online interface that presents the MEPI of a particular model under a particular reduction and asks the human evaluator to select the class for that MEPI from the list of twenty possible classes. The interface was shared again on a university forum and through social media. We first ran a control group with 25 trusted users, which logged an aggregate precision of 0.875 with a standard deviation of 0.061; in the open/online evaluation, we then filter user sessions more than two standard deviations from the control mean precision, giving a lower threshold of 0.753. The public evaluation then logged 531 valid user sessions according to the threshold, resulting in 11,588 valid classifications of machine MEPIs (equating to 26.2 classifications on average per session).

The results of the precision for cross-classification of MEPIs are summarised in Figure 2. Cells are shaded darker in order to visually indicate better performance. Darker rows indicate better precision for that classifier while darker columns indicate that the MEPIs produced by that classifier are easier for other classifiers. First we can confirm along the diagonal that all classifiers correctly predict (as expected by definition) all of their own MEPIs.

Machine results. With DNNs ordered by expected performance, we again see the clear trend that fewer reported classification errors in ILSVRC again correlate with better precision in the classification of MEPIs, with, for example, Se_{NET}Res_{NET}50 (SRN) having darker rows and lighter columns than other DNNs. We also see good cross-classifier performance for COLOUR and RESOLUTION: given that these are one dimensional reductions, the space of possible images is greatly reduced. On the other hand, DNNs struggle in cross-classification of the MEPIs under CROP and COMBINED: this is perhaps due to the larger search space for these reductions, but also indicates that these MEPIs are characteristic of the given DNN.

Human vs. machine results. Humans generally perform the best of all classifiers in this experiment, being adaptable enough to classify the MEPIs under all reductions with relatively high precision. The most difficult MEPIs for humans are in the CROP and COMBINED configurations for SRN, where a relatively high classification precision of 0.74–0.76 is still seen. We show some examples in Figure 3 of MEPIs that humans failed to classify correctly. With respect to the previously discussed ambiguity in the results of the previous sub-section, we can thus see that although CROP and COMBINED are more difficult cases for humans, issues with the more complex interface for these cases may explain the relatively poor results of humans in Figure 1 for CROP and COMBINED. Furthermore, we can see that DNNs often struggle to classify the human MEPIs, where the best classification precision reached was 0.43 for RN in the case of COLOUR. Figure 4 provides a sample of four human COMBINED MEPIs not correctly classified by any DNN model, illustrating the most difficult such cases for machine models.

6 DISCUSSION

Based on the presented experiments using our proposed frameworks for computing minimal-entropy positive images (MEPIs) in the case of both machine and human classifiers, we observe that:

- When compared with human-level performance on the same task, state-of-the-art machine models are sensitive (entropy-wise) to reductions in resolution when compared with other

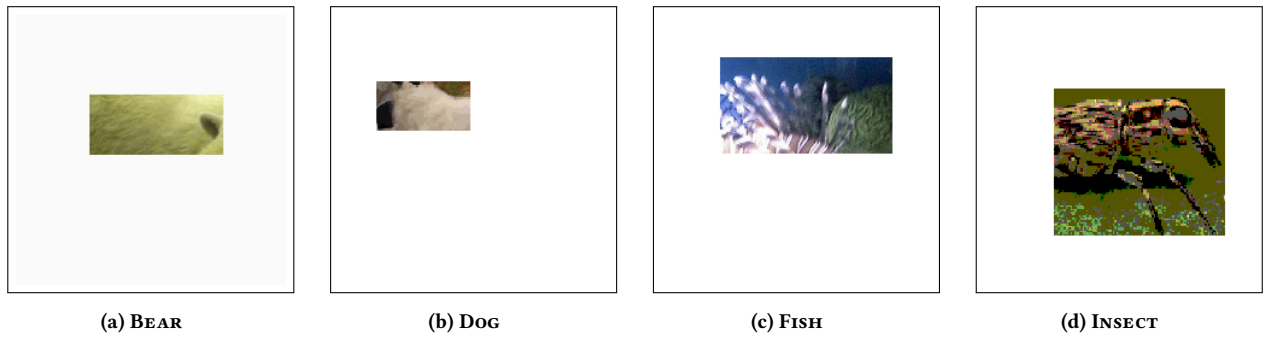


Figure 3: Examples of machine MEPIs not classified correctly by any human user

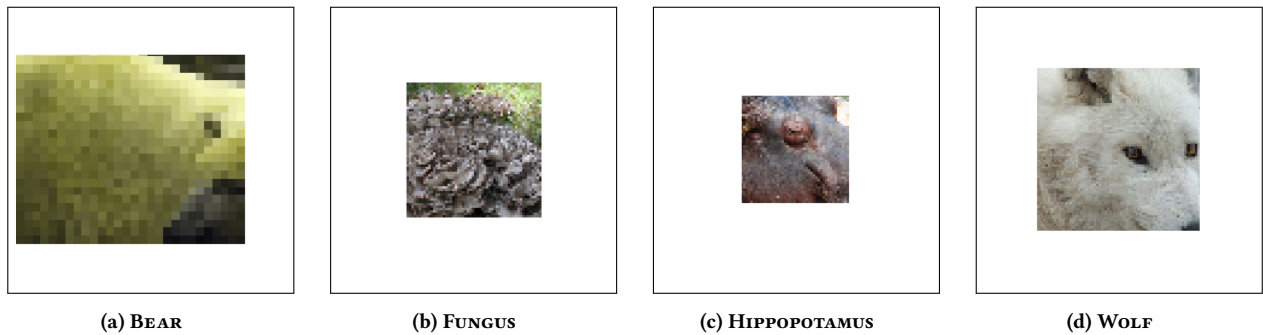


Figure 4: Examples of human MEPIs not classified correctly by any machine model

forms of reduction based on cropping or colour; such observations appear to independently support previous results indicating a bias towards texture in such models [6].

- Humans are relatively sensitive to the cropping of images, but are much less sensitive than machine classifiers to reductions in resolution; this tends to suggest that humans rely more on form and context for image classification, rather than textures of small regions of the image.
- In the cross-classification results, humans greatly outperform machine models for classifying the MEPIs of other models, suggesting a generally more robust aptitude in humans for the laconic classification of images.
- The machine models that perform better for laconic classification with respect to low-information images are those that perform better for traditional classification setting with respect to the full-quality images, suggesting that a possible way to improve traditional classification performance is to explore strategies for reducing the amount of input information that a model requires for correct classification.

Our work has a number of limitations that could be addressed for future work. While the method and interface used for computing human MEPIs in a bottom-up fashion work well for simpler (i.e., one-dimensional) forms of image enhancement, the unexpected result in Figure 1 showing an increase in the size of MEPIs in the combined case suggests that annotators had difficulty using the interface for multi-dimensional settings. Given that we filter incorrect images from consideration, we suspect that users “overshot” the

MEPI in one dimension, finding it difficult to select the particular dimension that is most likely to help them correctly classify the image. Another limitation is that for computing human MEPIs in the case of CROP and COMBINED, the void image begins with a central pixel, which may be far from the relevant region for classification. Refinements of the human MEPI framework would be interesting to explore in the future, for example to replace CROP with a more higher-level component-based analysis [31].

In order to facilitate the participation of non-expert humans, we selected 20 high-level classes. Within each class, we sampled 15 images. The limited number of classes and images anticipated the cost of human labelling in this work, where for comparing entropy ratios, we require users to manually generate MEPIs for 1,200 image–reduction combinations, while for cross-classification, given the four DNNs models considered, we require users to manually classify 4,800 image–reduction–model combinations. In future work, it would be interesting to develop datasets and results that consider more diverse classes, images, and models.

We have performed experiments for pre-trained, off-the-shelf DNN models. An interesting line of research would be to train models specifically for the task of laconic classification. Along these lines, one could consider computing MEPIs in the training set and feeding them (potentially recursively) back into the model; unlike standard data augmentation practices, such a method is guided by the classifier’s current performance. It would further be interesting to explore how models trained in such a manner perform in more traditional classification metrics – error rates,

precision, etc. – on the original input images, as well as whether or not they might help to address the observed lack of robustness of state-of-the-art DNNs in the presence of noisy [2, 3, 14, 22] or incomplete information [12, 19, 24, 27, 28, 31], or their lack of generalisation [7], or their bias towards texture [6]. It is important to note that the entropy of an image may increase as certain distortions of practical interest are intensified: more suitable general measures of robustness in such settings are left to be explored.

Laconic classification can help to understand and explain black-box classifiers. Looking at the MEPIs of each classifier provides insights into how they operate relative to other classifiers. Herein, for example, we have shown the gap between human and machine classification for low resolution or low colour images. Benchmarking tools for machine vision against human-level performance is an intuitive direction in which to gauge advances in the area. Laconic classification provides a new, general perspective from which such comparisons can be made. In this work we have considered DNN-based models as black boxes that we have tried to understand empirically. It would be interesting in future work to compare the sensitivity of different types of (DNN) architectures to different types of reductions, which may help to understand the effects of varying design choices on the classification process.

We publish various resources online to facilitate further research. Of particular interest are the human MEPIs that we have computed as part of this work, which can serve as a yardstick for human-level performance on the image classification task (see Figure 4 for examples). The best-performing pre-trained model achieves a precision of 0.285–0.424 on this dataset (depending on the type of reduction considered), where an interesting challenge for future work is to develop/train models that can surpass this precision.

Online Material. Further material for the paper can be found here: <http://aidanhogan.com/laconic/>.

ACKNOWLEDGMENTS

This work was supported by the Millennium Institute for Foundational Research on Data (IMFD). Hogan is also supported by Fondecyt Grant No. 1181896. Perez is also supported by Fondecyt Grant No. 1200967. We thank the users of our study and also thank the anonymous reviewers for their helpful feedback.

REFERENCES

- [1] Nilesh N. Dalvi, Pedro M. Domingos, Mausam, Sumit K. Sanghai, and Deepak Verma. 2004. Adversarial classification. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 99–108.
- [2] Samuel Dodge and Lina Karam. 2019. Human and DNN Classification Performance on Images With Quality Distortions: A Comparative Study. *ACM Transactions on Applied Perception (TAP)* 16, 2 (2019), 7:1–7:17.
- [3] Samuel F. Dodge and Lina J. Karam. 2017. A Study and Comparison of Human and Deep Learning Recognition Performance under Visual Distortions. In *International Conference on Computer Communication and Networks, (ICCCN)*. IEEE, 1–7.
- [4] Logan Engstrom, Brandon Tran, Dimitris Tsipras, Ludwig Schmidt, and Aleksander Madry. 2019. Exploring the Landscape of Spatial Robustness. In *International Conference on Machine Learning (ICML)*. PMLR, 1802–1811.
- [5] Alhussein Fawzi and Pascal Frossard. 2015. Manitest: Are classifiers really invariant?. In *British Machine Vision Conference (BMVC)*. BMVA Press, 106.1–106.13.
- [6] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. 2019. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations (ICLR)*. OpenReview.net.
- [7] Robert Geirhos, Carlos R. Medina Temme, Jonas Rauber, Heiko H. Schütt, Matthias Bethge, and Felix A. Wichmann. 2018. Generalisation in humans and deep neural networks. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*. 7549–7561.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE Computer Society, 1026–1034.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 770–778.
- [10] Olivier J. Hénaff and Eero P. Simoncelli. 2016. Geodesics of learned representations. In *International Conference on Learning Representations (ICLR)*.
- [11] Dan Hendrycks and Thomas G. Dietterich. 2019. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In *International Conference on Learning Representations (ICLR)*.
- [12] Tien Ho-Phuoc. 2018. CIFAR10 to Compare Visual Recognition Performance Between Deep Neural Networks and Humans. arXiv:1811.07270.
- [13] Hossein Hosseini and Radha Poovendran. 2018. Semantic Adversarial Examples. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*. IEEE Computer Society, 1614–1619.
- [14] Hossein Hosseini, Baicen Xiao, Mayoore Jaiswal, and Radha Poovendran. 2018. Assessing Shape Bias Property of Convolutional Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*. IEEE Computer Society, 1923–1931.
- [15] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. 2019. Squeeze-and-Excitation Networks. arXiv:1709.01507v4.
- [16] Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *CoRR* abs/1602.07360 (2016).
- [17] Andrei N. Kolmogorov. 1998. On Tables of Random Numbers. *Theor. Comput. Sci.* 207, 2 (1998), 387–395. [https://doi.org/10.1016/S0304-3975\(98\)00075-9](https://doi.org/10.1016/S0304-3975(98)00075-9)
- [18] Kieran G. Larkin. 2016. Reflections on Shannon Information: In search of a natural information-entropy for images. *CoRR* abs/1609.01117 (2016).
- [19] Drew Linsley, Sven Eberhardt, T. Sharma, P. Gupta, and Thomas Serre. 2017. What are the Visual Features Underlying Human Versus Machine Vision?. In *IEEE International Conference on Computer Vision (ICCV) Workshops*. IEEE Computer Society, 2706–2714.
- [20] Anh Mai Nguyen, Jason Yosinski, and Jeff Clune. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 427–436.
- [21] Michael J. D. Powell. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal* 7, 2 (1964), 155–162.
- [22] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.
- [23] Jie Shao, Xiaoteng Zhang, Zhengyan Ding, Yixin Zhao, Yanjun Chen, Jianying Zhou, Wenfei Wang, Lin Mei, and Chuanping Hu. 2016. Good Practices for Deep Feature Fusion. *ECCV 2016 Talk*.
- [24] Sanjana Srivastava, Guy Ben-Yosef, and Xavier Boix. 2019. Minimal Images in Deep Neural Networks: Fragile Object Recognition in Natural Images. In *International Conference on Learning Representations (ICLR)*. OpenReview.net.
- [25] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–9.
- [26] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus. 2014. Intriguing properties of neural networks. In *International Conference on Learning Representations (ICLR)*.
- [27] Shimon Ullman, Liav Assif, Ethan Fetaya, and Daniel Harari. 2016. Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences (PNAS)* 113, 10 (2016), 2744–2749.
- [28] Farahnaz Ahmed Wick, Michael L. Wick, and Marc Pomplun. 2016. Filling in the details: Perceiving from low fidelity images. arXiv:1604.04125.
- [29] Yue Wu, Yicong Zhou, George Saveriades, Sos S. Agaian, Joseph P. Noonan, and Premkumar Natarajan. 2013. Local Shannon entropy measure with statistical tests for image randomness. *Inf. Sci.* 222 (2013), 323–342.
- [30] Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, and Dawn Song. 2018. Spatially Transformed Adversarial Examples. In *International Conference on Learning Representations (ICLR)*. OpenReview.net.
- [31] Zijian Zhang, Jaspreet Singh, Ujwal Gadrijar, and Avishek Anand. 2019. Dissimilarity Between Human and Machine Understanding. *Proceedings of the ACM on Human-Computer Interaction (PACMHCI)* 3, CSCW (2019), 56:1–56:23.