

CHARACTERIZING IMAGE CLASSIFICATION DIFFICULTIES THROUGH REDUCED-DIMENSION CLASS CONVEX HULL ANALYSIS

Virtual Conference 19–23 October 2020

Shawn McGrory*, P. Michael Furlong[†], Krzysztof Skonieczny*

*Department of Electrical and Computer Engineering, Concordia University,
Montreal, QC, H3G1M8, Canada, E-mail: kskoniecz@encs.concordia.ca

[†]NASA Ames Intelligent Robotics Group (KBR, Inc), USA, E-mail: padraig.m.furlong@nasa.gov

ABSTRACT

The classification of natural terrain ahead of an autonomous rover can help it make decisions regarding the traversability of various paths, but remains an open research problem. Recent promise shown by deep learning algorithms in general image classification has seeded research aimed at transferring these successes to visual terrain classification. Despite the popularity of deep learning networks, there is still little work available on informed neural network design procedures for specific tasks such as terrain image classification. This work presents an algorithm that can be used to quantify the difficulty of specific image classification tasks and to investigate the characteristics of particular difficulties and trends in a way that is interpretable by humans. An accompanying analytical procedure characterizes such image classification difficulties; identifying what makes some images easily distinguishable exemplars of their class and what makes others readily confused with other classes.

1 INTRODUCTION

Autonomous robotics has seen considerable interest directed at transferring deep learning successes to their own applications - a task shown to demand substantial effort, especially for areas requiring novel architectures. Particularly relevant to this issue is the scarcity of knowledge regarding network design. As consequence, manual and automated network design methods necessitate the training or evaluation of many architectures, with the combined requirement frequently seen [1][2].

While there is still much to learn about deep learning algorithms, there is at least consensus that any finite network's performance may be constrained by its capacity [3]. This trait is not the sole arbiter of performance, however. For a given task, the amount of information which can be learned is related to the difficulty of the task, with more complex datasets commanding higher-capacity architectures. The capacity of a network may be explained as a quantity related to architecture design parameters, however,

producing a similarly concise explanation for dataset difficulty is less straightforward.

Image quality [4], object variations [5][6][7], and scene information [8][9], encompass some of the many properties observable within a dataset, potentially explaining confusing scenarios. Approaches more similar to our own are those proposing the measurement or identification of certain characteristics to predict the effort demanded by a recognition task. An example fitting this criterion is the work done by Russakovsky et al. [5] wherein authors developed a metric based on the number of possibly-object containing windows sampled before a true class-object is found. Similar instances of literature which explored difficulty prediction through measurement of image characteristics are [7] and [10] which considered difficulty as related to an various pixel value statistics. Although the consideration of identifiable traits may allow transfer of information to architecture design, methods thus far assume fixed difficulty attributes, not always guaranteed to be relevant.

Given that any predetermined feature is unlikely to capture all scenarios of image classification difficulty, some researchers have chosen to focus on complexity metrics with a more implicit relation to image characteristics. Recent work by Scheidegger et al. [11] has taken to this approach, determining dataset complexity as the performance achievable by various sized deep learning architectures. The authors' estimated difficulty, or effort, being a quantity found in relation to the networks' size and accuracy. Another example, taking a significantly different route, is Ionescu et al.'s study [6] of human response time as a predictor of image difficulty. In this study, authors define effort as the average time expended by human participants on a visual search task.

Summarizing - the use of interpretable image characteristics as a means to estimate classification difficulty has potential utility for architecture design, however current methods impose rigidly-defined features, thus limiting their scope. In contrast, approaches which are not explicitly concerned with attributes lending to challenge may possibly be adapted to a broader range

of use-cases, but in doing so, sacrifice the ability to extract information relevant to classifier design. While it can be argued that estimating difficulty using a method directly tied to neural networks might accomplish this goal, the resulting knowledge only informs *what* was successful, bearing no indication of *why* that was.

We attempt to consolidate the positive aspects of both noted directions, seeking to estimate dataset difficulty. In doing so, we achieve the following as primary contributions for our work:

- We propose a novel algorithm for sequencing images with respect to their difficulty, as defined by the size of their reduced representation necessary for class separation.
- We detail an analytical pipeline for extracting a dataset's image-class traits relevant to classifier design.
- We present case studies of our method applied to image datasets used for handwritten digit recognition, object recognition, and Martian terrain classification.

In the next section, we outline the ReDiHull algorithm, its main components, and its output data. Section 3 provides detailed instructions for producing and analyzing derived data to learn characteristics of a given image classification dataset. Section 4 both illustrates and discusses select results obtained by applying the algorithm and analysis for four different image classification datasets.

2 ALGORITHM

Our proposed algorithm is an iterative search for a reduced dimensionality representation where all classes in a dataset are contained in non-overlapping subspaces. In this section, we explain the ReDiHull (reduced-dimensionality class-convex hull) algorithm driving our method.

Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ represent a set of n samples $\{x_i\}_{i=0}^{n-1}$, produced by flattening images of equal height, width, and depth, into m -length vectors. We assume \mathbf{X} is normalized by subtracting the empirical mean $\bar{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i$ from each flattened image.

2.1 Dimensionality Reduction

Incremental analysis of the latent representation \mathbf{Z} produced for input data \mathbf{X} allows us to make inferential observations of both amount and type(s) of information required to separate classes. In order for interpretable results to be extracted, the dimensionality reducing model and resulting representation have the following requirements: (i) variables and transformed features are ordered by their optimality with respect to some metric and (ii) the change in distance

between any two points produced by a change in their representation's dimensionality $\mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$ must be greater than or equal to zero.

In order to apply the exact analytical procedure outlined in section 3 necessitates a third requirement: The chosen dimensionality reducing technique should also allow for visualization of the latent variables.

We will denote the transformation f which maps our input $\mathbf{X} \in \mathbb{R}^{n \times m}$ to a latent representation $\mathbf{Z}_d \in \mathbb{R}^{n \times d}$ as:

$$f(\mathbf{X}, d) : \mathbf{X} \rightarrow \mathbf{Z}_d \quad (1)$$

1) Principal Component Analysis: For this study, we achieve dimensionality reduction using principal component analysis (PCA) - an orthogonal, linear transformation, wherein samples are projected onto a coordinate system whose axes encode the directions of greatest variation within the applied training data. It follows from this definition that PCA satisfies the first imposed requirement by ranking principal components with respect to the amount of variation they capture. Further, compliance with the second requirement is guaranteed by the transformation's orthogonality.

2.2 Convex Hull

We model the class-occupied subspace as the convex hull produced by its constituent samples. Employing this definition allows us to evaluate our ability to distinguish between classes within a latent space, regardless of the corresponding transformation's relation to this criterion.

Whenever a convex-hull is referenced, we refer to its corresponding class as the *hull-class* ψ whose subset of dataset samples is given by \mathbf{X}^ψ . The convex-hull itself may be determined as the set of all convex combinations of points $x \in \mathbf{X}^\psi$, with n^ψ being the number of points that make up \mathbf{X}^ψ . As our analysis is performed exclusively on the latent representation $f(\mathbf{X}, d)$ we provide the convex-hull's definition as given by [12], written in terms of $z \in \mathbf{Z}_d^\psi$.

$$\text{Conv}(\mathbf{Z}_d^\psi) \equiv \left\{ \sum_{i=0}^{n^\psi-1} A_i z_i \mid \forall i : A_i \geq 0 \text{ and } \sum_{i=0}^{n^\psi-1} A_i = 1 \right\} \quad (2)$$

Now, with a class-occupied subspace modeled by that class' convex hull in the latent space, we determine overlap which we recall drives our definition of dataset complexity. With a *hull-class*' role defined, the next step it to form a *class-pair* (ω, ψ) for which to compute overlap, through selection of a *reference-class* ω . It follows from (2) that if a point $q \in \mathbf{Z}_d^\omega$

is not within the set of possible convex combinations produced by $\text{Conv}(\mathbf{Z}_d^\psi)$, it is not contained within the convex hull. We pose this as a linear programming problem of the following form:

$$\begin{aligned} \text{Find } q \text{ subject to } q &= \sum_{i=0}^{n^\psi-1} A_i z_i, \quad \sum_{i=0}^{n^\psi-1} A_i = 1 \\ \text{and } A_i &\geq 0 \text{ for all } i = 0, \dots, n^\psi - 1 \end{aligned} \quad (3)$$

This may be recognized as a linear feasibility problem, where q is a point belonging to the projected reference class ω and z_i is a point belonging to the projected hull class ψ .

2.3 Summary

The previous sections discussed the two central pieces of our ReDiHull algorithm - a dimensionality-reducing technique and class convex hull model. Here, we summarize the algorithm, allowing for understanding of its output.

As the dimensionality of the reduced-dimension representation is incrementally increased, between-class overlap can decrease. The number of dimensions needed to separate out the different classes is a measure of classification complexity.

ReDiHull is summarized in pseudo-code, shown as algorithm 1. We have as input (i) a reference-class subset, either training or testing, and (ii) a hull-class subset. We initialize the algorithm at dimension $d = 1$, with the set of overlapping points initialized with the entire reference-class subset, i.e., $\Theta_0^{(\psi\omega)} = \mathbf{X}^\omega$. The algorithm is presented in a form which produces full enumeration of overlaps at each d for a given class pair, by operating recursively. Expanding this to datasets having more classes amounts to applying ReDiHull to all possible class-pair permutations.

Algorithm 1: ReDiHull Algorithm

Input: $\Theta_{d-1}^{(\psi\omega)}, \mathbf{X}^\psi, d$
begin
 $\mathbf{Z}_d^\psi \leftarrow f(\mathbf{X}^\psi, d)$
 $\Theta_d^{(\psi\omega)} \leftarrow \{q \mid \forall q: q \subset \Theta_{d-1}^{(\psi\omega)} \text{ and } f(q, d) \in \text{Conv}(\mathbf{Z}_d^\psi)\}$
 if $|\Theta_d^{(\psi\omega)}| > 0$ **and** $d < m$ **then**
 ReDiHull($\Theta_d^{(\psi\omega)}, \mathbf{X}^\psi, d+1$)
 end
end

3 ANALYSIS

The ReDiHull algorithm, detailed in section 2, produces $\Theta = \{\Theta_1, \dots, \Theta_D\}$ the sets of points from

class ω which overlap with hull-class ψ in each examined latent space, where D is the largest d evaluated up to termination of the algorithm. This section outlines the procedure(s) applied to the algorithm's output, allowing us to uncover valuable characteristics hidden within the given dataset - characteristics such as those lending themselves to increased difficulty in distinguishing an ω -labelled image from the images of class ψ .

While the analysis will produce identical results when applied to output as is, we have found it to be conceptually advantageous to apply a simple transformation such that we instead examine the change in sets of overlapping points with respect to the change in dimensionality of the projective space. For a set of overlapping points at dimensionality $d-1$ and set of overlapping points at dimensionality d the change is simply the difference between the two sets:

$$\Xi_d = \Theta_d - \Theta_{d-1} \quad (4)$$

We refer to subset Ξ_d as the points which were *separated* at dimensionality- d . All visualized statistics, described in the sections 3.1 through 3.3, are obtained from the transformed algorithm output given by (4).

3.1 Locating Image Characteristics

Here we define the distributions representing the empirical probability of a point being separated at a given dimension. Following this definition, we explain the types of phenomena that we seek to identify within these distributions.

Now, let us consider a situation in which we are given Ξ_d the subset of test class ω 's points which were separated from the convex hull formed by class ψ using a d -dimensional space. This subset Ξ_d corresponds to a fraction of class ω 's total points and as such, this fraction is the empirical probability that a random point belonging to class ω , which overlaps with the convex hull formed by class ψ in a $d-1$ dimensional space, will be separated by incrementing to a d -dimensional space:

$$P(q \notin \text{Conv}(\mathbf{Z}^\psi) \mid d, q \subset \mathbf{Z}^\omega) = \frac{|\Xi_d|}{|\mathbf{X}^\omega|} \quad (5)$$

We note that $\bigcup_{d=1}^{d=D} \Xi_d = \mathbf{X}^\omega$.

By repeating this over each dimensional space for which overlap exists, we obtain an empirical separation probability distribution for a given class pair such as those in Fig. 3a. It is important to note that this probability distribution gives a measure of dataset complexity. Distributions with long tails extending to high values of d are capturing complexity that requires many dimensions to separate out.

Continuing in the context of Fig. 3a, we note an underlying primary distribution which is uni-modal

and spans the full range of dimensions shown. From our experiments, we have observed that this type of uni-modal shape occurs for the majority of image classes, albeit with slight variations.

In order to ascertain regions related to meaningful characteristics, we identify abnormal phenomena within the obtained distributions. Abnormality is determined relative to the primary distribution. An example of such an abnormality can be seen within Fig. 3a at the seventeenth dimension, where we observe an acute spike in separability, relative to the primary distribution. We have found that these types of spikes are particularly useful in determining *separating characteristics*; naturally-occurring characteristics which invoke a high degree of separability between respective image classes.

3.2 Determining Image Characteristics

In the previous step, we defined a separation probability distribution, then summarized trends and phenomena occurring within these distributions that may relate to potentially valuable dataset characteristics. The following step allows us to discover the attributes tied to the identified regions of interest. In order to achieve this goal, we construct projected value histograms from which we examine the latent model applied to samples forming the probability distribution.

For any determined probability distribution, there is a set of separated samples associated with each dimension, Ξ_d . Now we analyze the transformed values $f(\Xi_d)$ at each dimension. As an example, studying the transformed values might allow us to learn that a class' behavior at a certain dimensionality d is due to a strong, positive correlation with the k th latent variable, relative to the other classes. While this allows us to understand why the abnormality occurred, the supplementary information provided by visualizing the latent variables allow us to describe the image characteristic it relates to.

1) Description of the Projected Value Histograms:

Visualizing the distributions of projected values as a function of projected-space dimensionality is achieved through plots such as that shown in Fig. 3b.

Each plot corresponds to projections onto a single latent variable at index (k) indicated on the far-left. Within the referred figure, we are shown projections onto the latent variable at index (11). For this single latent feature, the maximum and minimum values, $\max \mathbf{Z}_{(k)}$ and $\min \mathbf{Z}_{(k)}$, are determined from the training set, then used to create representations of the positive and negative extremal observations, indicated by the tip of the upwards and downwards facing arrows respectively. The values may be seen by examining the chart's vertical axis as the maximum and minimum values are also used to define the

histogram. We partition the linear range formed by these extremal values into 40 bins of equal width. Application of a histogram allows for the 3D topology generated by the distribution of projected values to be visualized within a 2D space. Densely-distributed regions are shown using darker-colored bins whereas bins containing no elements are white. While each plot corresponds to projections onto a single latent feature, a plot's horizontal axis may extend to dimensionalities for which projected values $\mathbf{Z}_{(k)}$ have no contribution to separability. To distinguish between dimensionalities $d < k$ and $d \geq k$, histogram data is shown in grayscale for the former-defined range. Within the context of Fig. 3b, although we are shown projections relevant to the 17th component, we are shown projections shown for samples separated prior to the inclusion of this component, as well as those which precede it.

The black piecewise-linear trendline(s) show the change in mean of projected values between $d = k - 1$ and $d = k$ for the projection of separated images Ξ_d . The opacity of a curve segment is proportional to the number of points separated at $d = k$, with a higher opacity indicating more points separated. The standard deviations of projected values at each dimension are displayed as well, with downwards pointing chevrons denoting one standard deviation above the mean.

3.3 Visualizing Image Characteristics

When possible, we have found it beneficial to produce a visualization of the determined characteristic(s) using the images which form the abnormality or trend which was used to identify them. The qualitative information gained through these visualizations allows us to describe characteristic behaviors with greater certainty. Often, the combination of number of samples and dimensionality range for which the behavior is defined impedes our ability to extract useful information when all images are viewed simultaneously. To overcome this challenge, we substitute each dimension-specific image subset with a corresponding generalization.

1) *Description of the Image Difficulty Spectra:* Once again, the data forming these figures extends from the obtained class pair sets of separated samples associated with each dimension. To obtain a generalization \bar{x} of the images separated at each dimension, we simply compute the mean of the $n = |\Xi_d|$ number of images within each set $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Examples of these types of visualizations can be found at the bottom rows of Fig. 4 and Fig. 5a.

4 EXPERIMENTS & RESULTS

Here we present results in the form of case studies. The underlying intention is to (1) illustrate the outlined technique for improved clarity of its application,

(2) demonstrate concise examples of the types of extracted characteristics and their potential utility, and (3) validate the relevance of our method to the image classification task.

4.1 MNIST

Produced by [13], the MNIST dataset contains monochrome images of handwritten digits. For the context of this work, MNIST assumes the role of a less challenging image classification task, with even simple models achieving less than 10% error. To produce the discussed results, we designated 60 000 images to a training set and the remaining 10 000 to a test set.

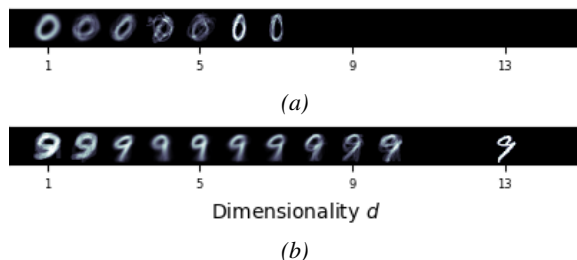


Figure 1: Difficulty spectra produced for MNIST (a) Images from digit Zero which are increasingly difficult to separate from digit One (b) Images from digit Nine which are increasingly difficult to separate from digit Seven.

The results shown in Fig. 1 provide examples of the visualizations described in 3.3, obtained through applying the algorithm to MNIST. These results were chosen as they illustrate clear visual interpretations of a trend-like behavior, as well as provide recognition scenarios of contrasting difficulty. For reference, a linear discriminant analysis (LDA) model trained on MNIST then evaluated against the test set never erroneously predicts an image containing digit Zero as containing digit One. When attempting to classify digit Nine images 50% of the error is attributed to digit Seven.

It is clear that for either scenario, images which require more dimensions to achieve separability appear increasingly similar to the respective convex hull-defining digit. Applying a more thorough examination reveals that the similarity can be attributed to distinct variations of the hull-defining digits. As example, the images of digit Zero which required 3 dimensions to separate from digit One are those which are most similar to the sans-serif, slanted variation of digit One.

Images in this dataset contain little to no variation in contrast, number of objects, or object's position(s). As such, using a metric which uses such attributes as a basis might provide indication that classification of MNIST images is not overtly difficult as a whole, however, such attributes would not sufficiently

describe the specific class-pair difficulties observed using our method. With the exception of serif differences, the digit variations lending themselves to increased difficulty are mostly linear transformations (e.g. rotation, scale, skew).

4.2 CIFAR-10

We apply our technique to CIFAR-10 - an image classification dataset depicting 10 classes of real-world objects and animals [14]. For results shown, the dataset was split into 50 000 training images and 10 000 test images. From this evaluation, we choose to highlight observations made in regards to the Deer class, then validate their relevance to classification through qualitative comparison of a feature computed using softmax regression, discriminating Deer from classes.



Figure 2: We trained a softmax classifier against CIFAR-10, then extracted the discriminating feature learned for Deer. This visual provides a reference for relating the observations drawn using our technique to a classification system.

We begin our discussion by examining the probability distribution shown in Fig. 3a. Doing so reveals a large spike in separability at seventeen dimensions, which, as mentioned in section 3.1, constitutes an anomalous behavior, with regard to the dominant trend. To determine the cause of the observed behavior, we refer to the corresponding projected value distribution (Fig. 3b). Through analysis of the projected data, it is clear that the characteristic related to this behavior is that images belonging to the Deer class have a stronger, positive correlation with the seventeenth loading vector, relative to images belonging to the Horse class. To the left of the projected value distribution, we are shown that the seventeenth loading vector encodes green (positive) to magenta (negative) overall image content. Summarizing these observations - images containing a significant amount of green content are likely to depict a Deer. Moreover, the result shown in Fig. 2 allows us to confirm this observation's relevance, demonstrating that softmax regression model determines the color green to be indicative of the Deer class.

One possible application would be define a simple feature detector, thus leveraging the separability observed for green-colored content. As mostly-green images usually correspond to Deer, a high-confidence

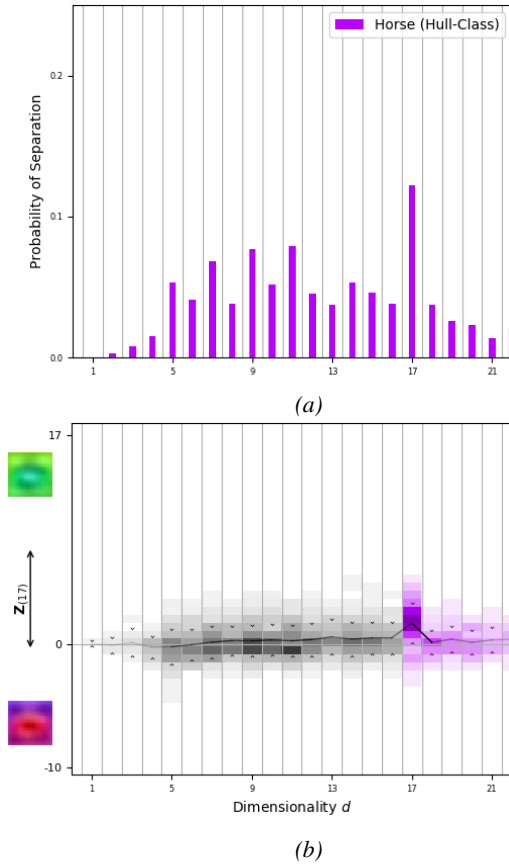


Figure 3: Resulting visualizations from applying our algorithm and analytical pipeline to CIFAR-10. (a) shows the probability of separation for a given dimensional-space when separating Deer from the Horse class convex hull. (b) shows the histograms of projected values, allowing for interpretation of the behavior observed in the probability distribution.

prediction could be quickly obtained for preliminary decision making, while simultaneously testing for additional features to refine the prediction. Moreover, the feature itself is mostly uniform thus it is likely that it may be compressed significantly.

If we assume some prior knowledge of the problem, the observation for separating Deer from other classes may lead one to formulate a converse application for this characteristic. Green colored background does not form a causal relation with Deer, thus preventing a classifier from associating this feature with Deer images may result in a more robust system.

4.3 JPL

The JPL dataset, procured by Shukla and Skonieczny in [15], serves as an example of a real-world scenario. Samples depict bedrock, sand, or rock-strewn terrains, selected from a subset of images captured by the Opportunity Mars Exploration Rover during its deployment. Due to the terrains' lack of structured-features and high visual similarity, the dataset contains many unique classification challenges.

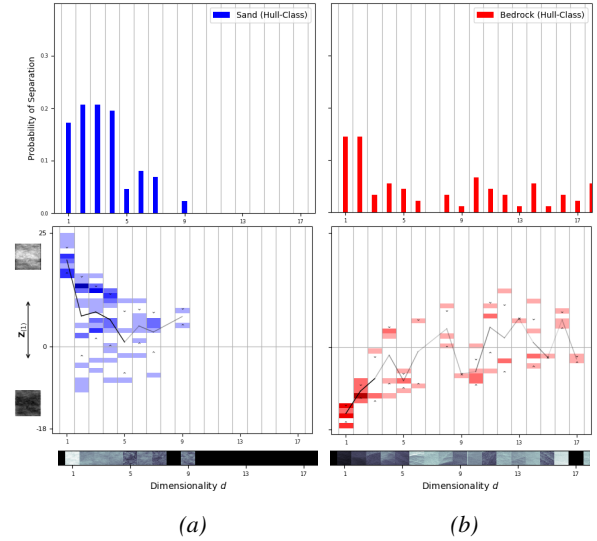


Figure 4: Results obtained for the Martian terrain dataset introduced by [15]. From the first to last row - probability of separation distributions, projected values histograms, and difficulty spectra for (a) Bedrock images separated from Sand (b) Sand images separated from Bedrock

Fig. 4 compares the statistics extracted when attempting to separate bedrock from sand's convex hull, to the opposite scenario of trying to separate sand from Bedrock's convex hull. In either corresponding probability distribution, there are no notable spikes of potential interest. Additionally, while bedrock displays a mostly, single modal probability distribution, no general behavior is apparent for the case of separating sand from bedrock. Moving on to the distributions of projected values onto the first loading vector, it is evident that Bedrock images requiring more dimensions to separate from sand have a projected value near zero, whereas those requiring few dimensions have a large, positive projected value. Conversely, Sand images which separate in few dimensions from Bedrock's convex hull have a large negative projection onto this loading vector. As sand images require more dimensions to separate, there is a clear trend that negative-valued projections tend towards zero.

By applying these explanations to the corresponding sets of mean images, our interpretations have a clear qualitative meaning - Bedrock images which are more difficult to distinguish from Sand are in general darker, which corresponds to difficulty arising during circumstances of low light or low exposure. In contrast, Sand images which are more difficult to distinguish from Bedrock are lighter overall, indicating classification difficulty when captured with high exposure.

4.4 CSA MET

Developed by Mission Control Space Services, the CSA MET dataset is composed of terrain images cap-

tured at the Canadian Space Agency Mars Emulation Terrain site in Quebec, Canada. This dataset shares a purpose with the JPL martian terrain dataset - training and evaluating autonomous soil assessment systems for autonomous path planning and risk assessment. The datasets differ in that the CSA MET samples are 3-channel (RGB) color images as opposed to the JPL datasets' grayscale NavCam images. Further, the task in the CSA MET dataset is semantic segmentation rather than image classification. Annotations are manually produced polygonal outlines of the relevant terrains within each image.

The goal of segmentation is to partition images such that each pixel is assigned a class label. To analyze segmentation datasets using our technique, we partition the the annotated images into smaller, square patches, providing sample images of unmixed terrain classes.

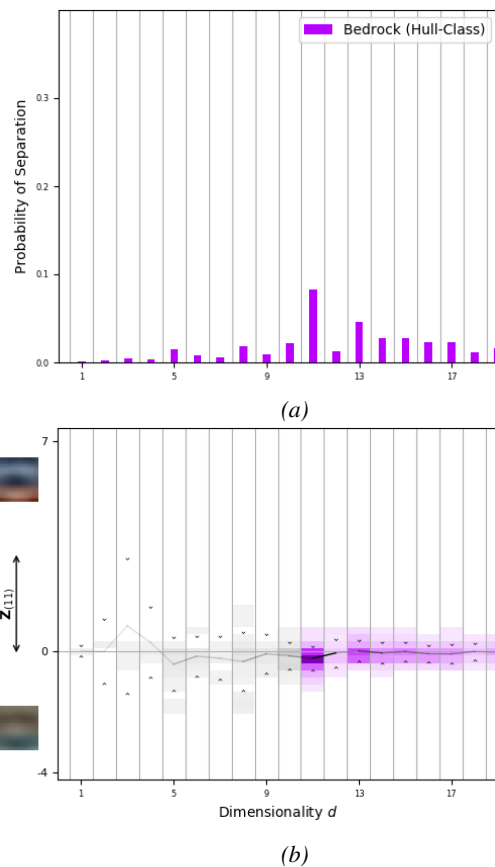


Figure 5: Resulting visualizations from applying our algorithm and analytical pipeline to the CSA MET dataset. (a) shows the probability of separation distribution for Sand images separated from Bedrock images. (b) shows the PCA projected value histograms for the 11th component using the CSA MET terrains Sand as the reference-class and Bedrock as the hull-class.

1) CSA MET - Anomalous Spike: The first observation to discuss is the presence of a large spike in

separability near the eleventh dimension when either *black sand*, *sand*, or *gravel* is applied as a reference-class and *bedrock* or *gravel* is applied as a hull-class. Fig. 5a. shows the separation probability distribution for separating sand from bedrock.

Recalling the discussion from section 3, we had noted that, in most situations, a spike in separability corresponds to the projection onto a distinguishing or separating feature.

Fig. 6 shows the first 11 PCA loading vectors computed for this dataset. Note that the 11th loading vector is the first to encode a color gradient, where part of the filter is blue and the other part is non-blue.

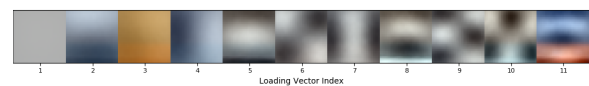


Figure 6: The top-11 PCA loading vectors computed for the CSA MET dataset.

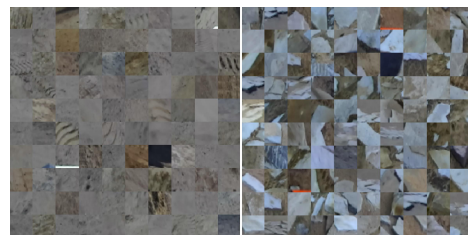


Figure 7: 100 samples each of color sand (left) and bedrock (right) images from the CSA MET dataset. Bedrock images include sub-regions of bluish stone and non-blue soil.

Fig. 7 shows 100 samples each of sand and bedrock images from the CSA MET dataset. It is clear that the bedrock class contains many images with distinct sub-regions (of bluish stone and non-blue soil), aligning with the main information that component 11 encodes. Sand images do not contain such color-distinct sub-regions.

To explicitly quantify the gradient in blueness between stone and soil sub-regions of bedrock images, as opposed to simply a gradient in brightness that could have been captured by loading vectors 2 and 4 for example, grayscale images of the bedrock are compared to the bedrock images' blue channel in Fig. 8. The blue channel data clearly shows higher contrast than the grayscale data; this higher contrast is further quantified by a higher standard deviation in pixel values for the blue channel ($\sigma = 45$) vs. grayscale ($\sigma = 35$).

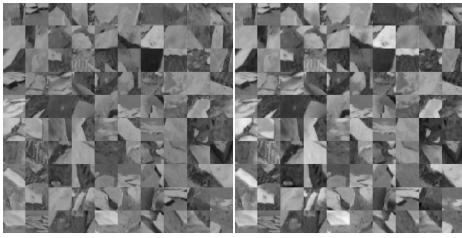


Figure 8: Grayscale (left) and blue channel (right) of CSA MET bedrock images. Blue channel shows higher contrast.

On the other hand, there is very little difference between grayscale and blue channel for sand (see Fig. 9; $\sigma = 17$ vs. $\sigma = 19$).

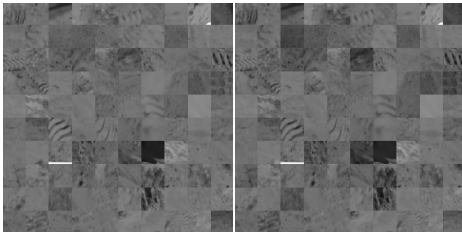


Figure 9: Grayscale (left) and blue channel (right) of CSA MET sand images, with little discernible difference in contrast.

Now looking at the projected value histogram for the 11th loading vector in Fig. 5b, the large spike in separations of sand from bedrock at component 11 has a projected value centered very near 0. Combining all the information discussed above suggests that sand images become distinguishable from bedrock, in the color CSA MET dataset, when it becomes clear the sand images *do not* include much of a color gradient.

5 CONCLUSIONS

We worked towards an informed approach to designing image classification systems, introducing an algorithm and corresponding analytical procedure for characterizing attributes related to classification difficulties. The proposed ReDiHull algorithm operates on incrementally ordered, reduced-dimension representations of a dataset, computing image-class overlap statistics within each latent-space. Data derived from the resulting statistics was used to produce probability distributions, projected value histograms, and image difficulty spectrum visualizations. Analysis of these allowed us to identify anomalies and trends, uncovering image characteristics related to the classification task. We applied the outlined method to several image datasets, demonstrating exemplars of inferred characteristics and their relation to image classification.

Acknowledgement

The authors acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) as well as funding and technical collaboration from Mission Control Space Services Inc.

References

- [1] A. Zela, A. Klein, S. Falkner, and F. Hutter, "Towards automated deep learning: Efficient joint neural architecture and hyperparameter search," *CoRR*, vol. abs/1807.06906, 2018. [Online]. Available: <http://arxiv.org/abs/1807.06906>
- [2] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *CoRR*, vol. abs/1707.07012, 2017. [Online]. Available: <http://arxiv.org/abs/1707.07012>
- [3] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [4] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Computer Vision – ECCV 2008*, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 386–399.
- [5] O. Russakovsky, J. Deng, Z. Huang, A. C. Berg, and L. Fei-Fei, "Detecting avocados to zucchinis: What have we done, and where are we going?" in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 2064–2071.
- [6] R. Tudor Ionescu, B. Alexe, M. Leordeanu, M. Popescu, D. P. Papadopoulos, and V. Ferrari, "How hard can it be? estimating the difficulty of visual search in an image," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [7] D. Liu, Y. Xiong, K. Pulli, and L. Shapiro, "Estimating image segmentation difficulty," vol. 6871, 08 2011, pp. 484–495.
- [8] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," vol. 1, 11 2005, pp. 90 – 97 Vol. 1.
- [9] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," 11 2009, pp. 2232 – 2239.
- [10] S. Vijayanarasimhan and K. Grauman, "What's it going to cost you?: Predicting effort vs. informativeness for multi-label image annotations," 01 2009, pp. 2262–2269.
- [11] F. Scheidegger, R. Istrate, G. Mariani, L. Benini, C. Bekas, and A. C. I. Malossi, "Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy," *CoRR*, vol. abs/1803.09588, 2018. [Online]. Available: <http://arxiv.org/abs/1803.09588>
- [12] R. T. Rockafellar, *Convex analysis*, ser. Princeton Mathematical Series. Princeton, N. J.: Princeton University Press, 1970.
- [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278 – 2324, 12 1998.
- [14] A. Krizhevsky, "Learning multiple layers of features from tiny images," *University of Toronto*, 05 2012.
- [15] D. K. Shukla and K. Skonieczny, "Simple texture descriptors for classifying monochrome planetary rover terrains," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 5495–5502.