https://doi.org/10.2352/issn.2168-3204.2022.19.1.07

Enhanced Computer Vision Using Automated Optimized Neural Network Image Pre-Processing

Kevin Fenton, Colorado State University, Ft. Collins, CO USA
Vincil Bishop, Colorado State University, Ft. Collins, CO USA
Steven J. Simske, Colorado State University, Ft. Collins, CO USA

Abstract

This research focuses on the benefits of computer vision enhancement through use of an image pre-processing optimization algorithm in which numerous variations of prevalent image modification tools are applied independently and in combination to specific sets of images. The class with the highest returned precision score is then assigned to the feature, often improving upon both the number of features captured and the precision values. Various transformations such as embossing, sharpening, contrast adjustment, etc. can bring to the forefront and reveal feature edge lines previously not capturable by neural networks, allowing potential increases in overall system accuracy beyond typical manual image pre-processing. Similar to how neural networks determine accuracy among numerous feature characteristics, the enhanced neural network will determine the highest classification confidence among unaltered original images and their permutations run through numerous pre-processing and enhancement techniques.

Motivation

The motivation came from an image processing course when experimenting on how various image transformations impacted computer vision; specifically, the number of features captured and the levels of confidence in the image classification. Even with high performance image recognition tools such as GoogleVision or AlexNet [1], the levels of confidence could be improved upon and the number of captured features increased with various image transformations.

Problem

Current image recognition applications do not take an automated approach of applying varying image transformations to improve upon returned features and levels of confidence [2]. Manual optimization is largely impractical due to the exponential number of image modifications and combinations possible. Even minor image alterations, however, can lead to significant changes in how computer vision interprets a desired image. An automated solution allows for an assessment of each image from various perspectives allowing the neural network to benefit from the most advantageous transformation combination [3].

Approach

The ideal system would include algorithms that would automatically run each potential combination of *n* filters selected for input alterations (see Figure 1). A cloud vision API was used for classification and assessing which pre-processed method was more effective in maximizing overall accuracy [4]. The neural network was trained with images of the desired features to be identified and classified. For our example, we used the Google Cloud Vision API [5], a large-scale neural network which offered pre-trained learning models and classifications into millions of predefined categories. The Google Cloud Vision was used to determine the classification and percentage likelihood that the image has been correctly classified.

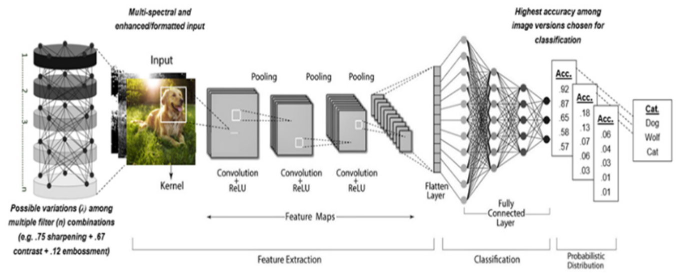

***Figure 1**. Enhanced Neural Network Process* [6]

The inherent algorithms determined feature characteristics such as edges, boundaries, locations, and number pixels that were used to measure this degree of confidence of classification [7]. Each image transformation brought about variations in these characteristics. Each differently pre-processed image contained its own unique characteristics that were measured; and through this input layer, predictions and comparisons were made to the trained images [8]. Comparisons of the feature characteristics were made to the millions of available images accessible through the Google Cloud Vision API.

The experiment was performed in two parts. The first was a preliminary experiment to gauge the impact various transformations had on image recognition accuracy and determine which transformations would be better candidates for inclusion in a more robust automated approach. The preliminary experiment was done using 140 combinations of various common filters. For this experiment, sample photos were altered with various combinations of image transformations using ImageJ and run through the GoogleVision API. In addition to the original unaltered images, initial transformations used were edge detection, embossment, inversion, sharpness, contrast, background subtraction, local contrast, and despeckling.

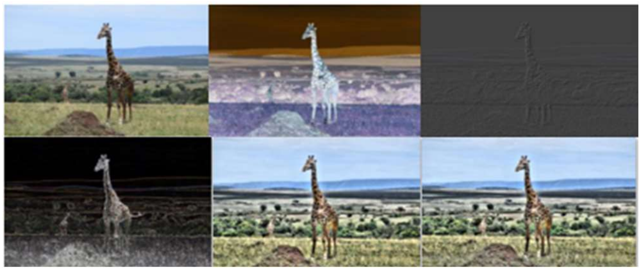

Figure 2. Preliminary tests included the Original (first image), Invert (0.95 accuracy), Embossment (0.00), Edge Detection (0.80), Local Contrast (0.76), and Local Contrast + Sharpness (0.95)

Based upon the preliminary results, an automated solution was then developed in which sample photos were altered with various image transformations using the open source ImageMagick [9] and BackgroundRemover [10] software libraries. Node.js [11] was used to automate the process and PostgreSQL [12] was used to collect and relate the results. Given the time allocation needs for establishing a functional auto-optimization algorithm and tool, the approach for this experiment was to simply simulate the benefits on a smaller scale by showing how variations of image alterations can impact image classification confidences. 15 original images were chosen to provide a variety of different perspectives and image attributes.

The original subject images were all taken of animals in a natural/safari type setting. These original images and more relevant artifacts can all be found in the Computer Vision Reproduction Package found at the end of this document. For the image modification portion of the experiment, four operation types (sharpen, contrast, despeckle, and background removal) were chosen to influence computer vision accuracy. The sharpen and contrast modifications were varied at four intervals: 25%, 50%, 75%, and 100%.

The software used for the experiment did not allow for the variance of the despeckle and background removal operations, so these were both applied at 100% where used. Solitary modifications were applied to each subject image (e.g., sharpen @ 25% only or contrast @ 50% only) as well as dual combinations of each modification operation. For operations applying dual modifications (e.g., sharpen @ 25% and contrast @ 50%) 66 unique image modification combinations of all operation permutations were applied in pairs to each original image, except for "like operations". For example, the contrast operation at 100% would not also be applied to an image that was already modified with the same contrast operation at 100%. The order in which modifications are applied affects identification results, so modification combinations with like operations in a different order were treated as unique. ***The combination of all operation permutations generated a total of 1,155 total samples*** [13]. ***This included 15 original unmodified images, 150 permutations with a single modification and 990 permutations with a combination of two modifications*** that were analyzed by the neural network [14].

Results from the computer vision identification operation will come from label and object detection features of the Google Vision API. Each identified object will contain a string describing the object as well as a decimal score indicating how confident the algorithm was in identifying the object correctly (confidence score). To measure the effects of different image modification combinations on object detection, the confidence score of all objects detected in the image will be averaged. The calculation does not consider whether the identified objects changed or not between image modification combinations, the possibility of this examination has been reserved as a future research opportunity.

The algorithm created for this research determines the maximum Neural Network accuracy (**NNacc**) for all image pre-processing filters used by the system for all image pre-processing filters.

$$\boldsymbol{maxNNacc}(\boldsymbol{\sigma_{1:n}}: \mathfrak{S}(\sigma_n))$$

Where

$\boldsymbol{\sigma_{1:n}}$ = the number of filters used/transformations performed on any given image *(e.g., despeckle, background substitution, contrast, sharpening, etc.)*.

$\mathfrak{S}(\sigma_n)$ = all permutations of these various filter combinations (*e.g., 22% filter combined with 13% filter 2 combined with 100% filter 3*).

For simplification in our dataset, we used 25% filter increments for the various combinations. As the permutations are processed, a log is maintained of the highest topicality for the desired subject(s) in the image and the number of desired features captured. As new combinations with higher accuracy rates are discovered, lower performance image alterations can be discarded.

Results

The various imaging operations in combination with CNN approaches added features not generally associated with the combination of original images + CNN approaches. This yielded computer vision resilience in areas not currently addressable, such as remote sensing areas with limited input data sets. Essentially, the images and their transforms created a larger data set, with the transformed images operating as "simulated images" in some cases to increase the effective level of training.

In the preliminary experiment, of the 140 various images used, the original only recorded the highest level of confidence on 3 occasions proving the need for an automated transformation approach. Nearly each of the combinations had significant impacts on confidence for the different images used. For instance, the sharpen filter moved pixels away from their clustered values to add definition and enhance edges reflecting an increase in entropy in the image. Despeckle did the opposite as it removed noise and decreased entropy. This was beneficial for images where the decrease in classification confidence was less about defined edges and more about excess background noise that was impacting the ability to clearly classify the primary subjects.

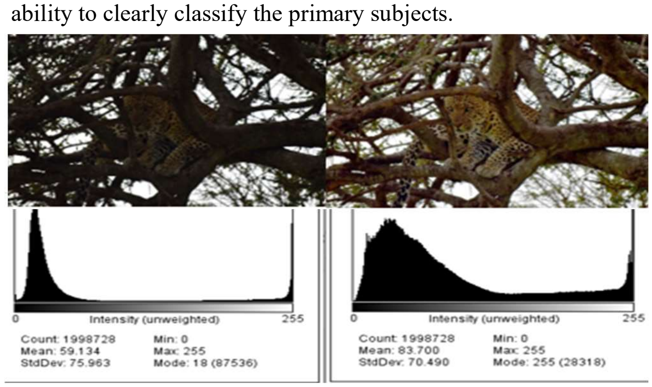

Figure 3. Contrast and Despeckling and Background Subtraction Yielded a 31% confidence improvement over the original image

While the edge detection and embossment filters performed relatively poorly overall in terms of the level of confidence, both did add to the number of subjects appropriately classified for the image that had numerous subjects in the background. The results proved that there was no single optimal filter that can be used universally for image classification, and showed the need for an optimization algorithm. Even with a limited data set, 4 different filters or combinations each produced optimal results for at least one image. Given Google Vision's immense training set, a 5% average increase in confidence was significant and worthy of further research.

The automated approach, while significantly expanding the efficiency and number of permutations analyzed, yielded the following results:

Original Images Results

Of the 15 original images, the mean score ranged from 0.7314 to 0.9273. The mean score of the sets was 0.8517 with a standard deviation of 0.0492, and the median score was 0.8602.

croc-1.JPG	0.9273
giraffe-1.JPG	0.9092
elephants-1.JPG	0.8895
animals-1.JPG	0.8842
cheetahs-1.JPG	0.8780
giraffes-2.JPG	0.8753
zebras-1.JPG	0.8678
lions-1.JPG	0.8602
lion-3.JPG	0.8494
converted.JPG	0.8436
hippo-2.JPG	0.8368
bonobo-1.JPG	0.8152
jaguar-1.JPG	0.8089
lions-2.JPG	0.7985
hippo-1.JPG	0.7314

***Figure 4**. Tableau output of original image confidence scores*

Modified Image Results

When the 15 original images were altered across all 1,140 modification combinations the average of scores per image range from 0.76308 to 0.86193. While many transformations outperformed the originals in terms of accuracy, some yielded significantly lower scores and lowered the overall average among these transformations. The set's mean score was 0.81817 and the median score was 0.82324 with a standard deviation of 0.03069.

giraffe-1.JPG	0.86193
elephants-1.JPG	0.85989
animals-1.JPG	0.84281
converted.JPG	0.84064
lion-3.JPG	0.83785
giraffes-2.JPG	0.83439
lions-1.JPG	0.82653
zebras-1.JPG	0.82324
croc-1.JPG	0.82282
cheetahs-1.JPG	0.81223
hippo-2.JPG	0.80511
lions-2.JPG	0.79042
bonobo-1.JPG	0.78029
hippo-1.JPG	0.77125
jaguar-1.JPG	0.76308

***Figure 5**. Tableau output of image transformations (and combinations) averages.*

Image Modification Combination Results

Of the 77 analyzed combinations (76 modification combinations with the original unmodified images treated as a control) the average of scores for each modification combination ranges from 0.7042 to 0.8607. The set's mean score was 0.8186 and the median score was 0.8456 with a standard deviation of 0.0485. The modification combination with the highest score on average was "contrast @ 50%" and the combination with the lowest score was "bgsubtract @ 100% + contrast @ 100%".

contrast @ 50%	0.860660
contrast @ 25%	0.860245
sharpen @ 75% + contrast @ 25%	0.859652
sharpen @ 50% + contrast @ 25%	0.859652
sharpen @ 25% + contrast @ 25%	0.859652
sharpen @ 100% + contrast @ 25%	0.859652
despeckle @ 100% + contrast @ 50%	0.859074
contrast @ 25% + despeckle @ 100%	0.858779
contrast @ 50% + sharpen @ 75%	0.857047
contrast @ 50% + sharpen @ 50%	0.857047
contrast @ 50% + sharpen @ 25%	0.857047
contrast @ 50% + sharpen @ 100%	0.857047
despeckle @ 100% + contrast @ 25%	0.856887
sharpen @ 25% + despeckle @ 100%	0.855962
sharpen @ 100% + despeckle @ 100%	0.855962

***Figure 6**. Tableau output with highest overall mean scores among the originals and all permutations (truncated to top 15 results)*

Results Summary

The overall results yield interesting relationships between image modification combinations, images, and their object identification scores. The statistics reported in these results seem to be specific to this data set. Given a different set of images, it is likely that the leading image modification combinations would be different depending on the constitution of the original images. In terms of highest accuracy, some variation of transformation (often differing from image to image) outperformed the original. The sequence of transformations also factored. Even when the same percentage of two transformations (e.g., 25% contrast 50% sharpness vs.50% sharpness, 25% contrast) occurred, accuracy rates different depending which transformation occurred first. Most notably, however, is that the data validated our hypothesis that an optimization algorithm is needed to obtain maximum accuracy given that no single transformation can obtain this alone.

Analysis

Top Image Modifications

animals-1.JPG	sharpen @ 25%	0.92523
	sharpen @ 50%	0.92523
	sharpen @ 75%	0.92523
	sharpen @ 100%	0.92523
bonobo-1.JPG	despeckle @ 100% + contrast @ 75%	0.85408
cheetahs-1.JPG	contrast @ 25%	0.88050
converted.JPG	contrast @ 50%	0.89039
croc-1.JPG	despeckle @ 100%	0.93028
elephants-1.JPG	sharpen @ 25% + despeckle @ 100%	0.89861
	sharpen @ 50% + despeckle @ 100%	0.89861
	sharpen @ 75% + despeckle @ 100%	0.89861
	sharpen @ 100% + despeckle @ 100%	0.89861
giraffe-1.JPG	contrast @ 25%	0.92153
giraffes-2.JPG	contrast @ 50%	0.91907
hippo-1.JPG	sharpen @ 25% + contrast @ 100%	0.87190
	sharpen @ 50% + contrast @ 100%	0.87190
	sharpen @ 75% + contrast @ 100%	0.87190
	sharpen @ 100% + contrast @ 100%	0.87190
hippo-2.JPG	contrast @ 75% + despeckle @ 100%	0.89223
jaguar-1.JPG	contrast @ 25% + despeckle @ 100%	0.84371
lion-3.JPG	contrast @ 50% + bgsubtract @ 100%	0.88932
lions-1.JPG	despeckle @ 100%	0.86683
lions-2.JPG	contrast @ 75%	0.85917
zebras-1.JPG	contrast @ 50% + sharpen @ 25%	0.89036
	contrast @ 50% + sharpen @ 50%	0.89036
	contrast @ 50% + sharpen @ 75%	0.89036
	contrast @ 50% + sharpen @ 100%	0.89036

***Figure 7**. Tableau output of top performing transformations (including original images) by image*

When modification combinations with top detection scores are examined on a per image basis, it was observed that three modifications yielded the highest average detection scores across two images, these modifications were: despeckle @ 100%, contrast @ 50%, and contrast @ 25%. The remaining images all had unique top modifications. In several cases certain modifications tied for average detection score. The observation that in most cases a different modification combination yielded top detection scores indicates that the modification that would yield the greatest improvement in detection score is likely dependent on qualities of the individual images themselves. Analysis of what qualitied might respond best to which modifications is reserved for a further research opportunity.

Improvements Over Original Images

- There were 69 image modifications that yielded detection score improvements over the original image detection score.
- These improvement percentages on average ranged from 0.6% to 17.5%.
- The set's mean improvement was 4.4% while the median improvement was 2.7% with a standard deviation of 0.0480.
- The image modification yielding the greatest improvement overall was "sharpen @ 75% + contrast @ 100%".

sharpen @ 75% + contrast @ 100%	17.5%
sharpen @ 50% + contrast @ 100%	17.5%
sharpen @ 25% + contrast @ 100%	17.5%
sharpen @ 100% + contrast @ 100%	17.5%
despeckle @ 100% + contrast @ 100%	17.1%
contrast @ 100%	15.0%
contrast @ 100% + despeckle @ 100%	12.8%
contrast @ 100% + sharpen @ 75%	12.0%
contrast @ 100% + sharpen @ 50%	12.0%
contrast @ 100% + sharpen @ 25%	12.0%
contrast @ 100% + sharpen @ 100%	12.0%
contrast @ 75% + despeckle @ 100%	4.5%
sharpen @ 75% + contrast @ 75%	4.3%
sharpen @ 50% + contrast @ 75%	4.3%
sharpen @ 100% + contrast @ 75%	4.3%

***Figure 8**. Tableau output of the difference between the highest performing transformation and the original*

Opportunities for Further Research

- What qualities of an image cause which image modification combinations to be more effective at increasing object identification confidence scores?
- Do image modification combinations equally or proportionally affect objects regarding their topicality score or might different image modification combinations yield differing/better results per object depending on their topicality or other factors?
- To what extent do image modification combinations affect the inventory of objects identified? Are there image modification combinations that can increase not only the confidence with which objects are identified, but also the number of objects identified?

Conclusion

The implications of this research could yield significant improvements to computer vision operations by increasing system levels of confidence, increasing the number of useful image features, and providing simulated images to improve training breadth, resilience, and robustness. At a minimum, the research provides insight into optimum combinations of various image transformation methods for image pre-processing prior to neural network classification operations. Continued areas of interest with this research include an analysis of the correlation of image entropy to the accuracy rate of the various combinations of images and continuing research into the additional transformation tools and their level of success for potential inclusion in such a system.

Reproduction Package

To aide in the reproduction and further analysis of the findings presented in this paper, a reproduction package has been made available online. The reproduction package contains all source code, raw data sets, as well as notes and other images that may be of interest to the reader. Please find this package online at this location: https://github.com/vincilbishop/image-recognition

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[2] B. Christopher, *Pattern Recognition and Machine Learning*. Cambridge, United Kingdom: Springer, 2006.

[3] R. Wang, W. Li, and L. Zhang, "Blur image identification with ensemble convolution neural networks," *Signal Processing*, vol. 155, pp. 73–82, Feb. 2019, doi: 10.1016/j.sigpro.2018.09.027.

[4] S. Li *et al.*, "All-optical image identification with programmable matrix transformation," *Optics Express*, vol. 29, no. 17, p. 26474, Aug. 2021, doi: 10.1364/OE.430281.

[5] Google, "Google Cloud Vision API." https://cloud.google.com/vision, 2022.

[6] Developers Breach, "Convolutional Neural Network | Deep Learning" , https://developersbreach.com, 2021

[7] Y. Liu, Y. Huang, J. Zhang, and H. Shen, "Anti-noise image source identification," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 19, Oct. 2019, doi: 10.1002/cpe.5104.

[8] R. Das, S. Thepade, and S. Ghosh, "Framework for Content-Based Image Identification with Standardized Multiview Features," *ETRI Journal*, vol. 38, no. 1, pp. 174–184, Feb. 2016, doi: 10.4218/etrij.16.0115.0102.

[9] ImageMagick Studio LLC, "ImageMagick." https://imagemagick.org/, 2022.

[10] J. Nader, "BackgroundRemover." https://backgroundremover.app/, 2022.

[11] OpenJS Foundation, "Node.js." https://nodejs.org/, 2022.

[12] The PostgreSQL Global Development Group, "PostgreSQL." https://www.postgresql.org/, 2022.

[13] V. Bishop and K. Fenton, "Computer Vision Reproduction Package." https://github.com/vincilbishop/image-recognition, 2022.

[14] N. Shabairou, E. Cohen, O. Wagner, D. Malka, and Z. Zalevsky, "Color image identification and reconstruction using artificial neural networks on multimode fiber images: towards an all-optical design," *Optics Letters*, vol. 43, no. 22, p. 5603, Nov. 2018, doi: 10.1364/OL.43.005603.

Author Biography

Kevin Fenton is a Ph.D. candidate in the Systems Engineering Department at Colorado State University, Ft. Collins. Kevin is an environmental systems engineer developing environmental software solutions for the Department of Defense.

Vincil Bishop is a Ph.D. student in the Systems Engineering Department at Colorado State University, Ft. Collins. Vincil is a software engineer by trade and enjoys developing methods that provide insight into contributor behavior.

Dr. Steve Simske is a professor of Systems Engineering at Colorado State University, Ft. Collins. Dr. Simske was at HP from 1994-2018, and was an HP Fellow, Vice President, and Director in HP Labs.