Neural Pixel Error Detection

Erika Varis Doggett
Anna M.C. Wolak
erika.varis.doggett@disney.com
anna.x.wolak.-nd@disney.com
Walt Disney Studio Technology
Burbank, California

P. Daphne Tsatsoulis
Nicholas McCarthy
daphne.tsatsoulis@accenture.com
nicholas.mccarthy@accenture.com
Accenture Labs
Dublin, Ireland

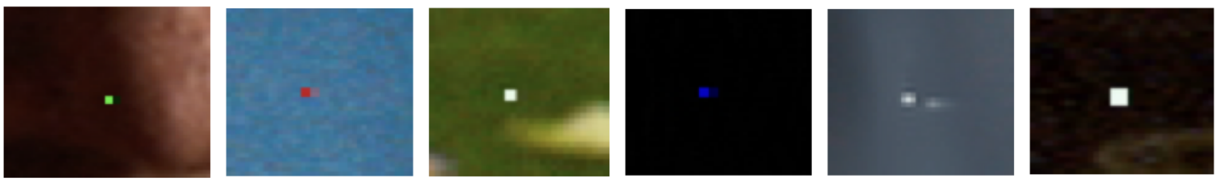

Figure 1: Sample created and natural pixel errors, property of The Walt Disney Company

ABSTRACT

Current video quality control entails a manual review of every frame for every video for pixel errors. A pixel error is a single or small group of anomalous pixels displaying incorrect colors, arising from multiple sources in the video production pipeline. The detection process is difficult, time consuming, and rife with human error. In this work, we present a novel approach for automated pixel error detection, applying simple machine learning techniques to great effect. We use an autoencoder architecture followed by statistical post-processing to catch all tested live action pixel anomalies while keeping the false positive rate to a minimum. We discuss previous dead pixel detection methods in image processing, and compare to other machine learning approaches.

CCS CONCEPTS

• **Computing Methodologies → Computer vision Tasks**; • **Computing methodologies** → *Anomaly detection*; *Neural networks.*

KEYWORDS

neural networks, autoencoder, anomaly detection, pixel errors, quality control, video

ACM Reference Format:
Erika Varis Doggett, Anna M.C. Wolak, P. Daphne Tsatsoulis, and Nicholas McCarthy. 2019. Neural Pixel Error Detection. In *Proceedings of SIGGRAPH '19 Talks.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3306307.3328197

SIGGRAPH '19 Talks, July 28 - August 01, 2019, Los Angeles, CA, USA

ACM ISBN 978-1-4503-6317-4/19/07.
https://doi.org/10.1145/3306307.3328197

1 INTRODUCTION

In the current video production pipeline, a studio often has a final quality control check for video errors at the pixel level before the asset is sent out for distribution. A pixel error is defined as a single or small group of pixels that displays the wrong color, and can arise from multiple sources in the video production pipeline. The color and temporal persistence of the errors may vary widely, and therefore are not easily identified by heuristic rules. In addition, some pixel errors may be visually indistinguishable from purposeful creative decisions. Thus, automated tooling must always allow human operators to verify the detected errors.

To address this problem in the video production pipeline, we propose a Neural Pixel Error Detection (NPED) system utilizing a neural network autoencoder in addition to temporal comparisons of pixel patches from frame to frame.

2 PREVIOUS APPROACHES

Previous solutions to this problem include image processing techniques such as Fourier and median filters to identify areas of high change from original to filtered images, with user-defined thresholds of change. When testing videos against such techniques, we found this heuristic frame-by-frame approach did not scale from images to video, and created far too many false positives.

Other algorithmic approaches in anomaly detection include kernel density estimation [Latecki et al. 2007] and variational autoencoders [An and Cho 2015; Rumelhart et al. 1986; Xu et al. 2018]. Initial tests utilizing KDE proved too computationally expensive on the evaluation side, making it impractical for our purposes.

3 METHOD

Since pixels are low semantic objects, standard deep learning vision models will learn dot filters, and more information is needed, such as temporality. Inspired by [An and Cho 2015], we take an autoencoder

Table 1: Recall & False Positive Frame Rates

Test Video	AE-10x10		AE-15x15		AE-20x20		CNN-temp	
	recall	FP rate	recall	FP rate	recall	FP rate	recall	FP rate
13th Warrior	✓	4.3%	✓	16.9%	✗	20.8%	✓	24.7%
Bedtime Stories	✓	12.7%	✗	21.0%	✗	45.2%	✓	29.4%
20,000 Leagues	✓	11.5%	✓	31.1%	✓	37.6%	✓	5.2%
Huckleberry Finn	✓	2.2%	✓	7.0%	✗	15.6%	✓	18.7%
Shopgirl	✓	19.9%	✓	35.1%	✓	38.9%	✓	44.3%
Rogue One	✓	1.4%	✓	15.1%	✓	20.6%	✓	21.3%
Averages	100%	8.67%	83.33%	21.03%	50%	29.78%	100%	23.93%

approach, beginning with a simpler network, and add temporal novelty detection.

We train a simple autoencoder on clean image patches and then use reconstruction errors to identify potential anomalies, paired with temporal novelty detection to reduce false positives. We tested models with different patch sizes, and compared to a CNN classification approach.

3.1 Data

For our training data, we compiled short non-anomalous video segments from a variety of styles and colors. Videos were segmented into frames and letterboxing removed. We created datasets for 10x10, 15x15, and 20x20 pixel patches to compare models.

For testing, we collected 6 live action video snippets that included discovered and generated pixel errors; each test video ranged from 438 to 700 frames. Titles include *The 13th Warrior*, *Bedtime Stories*, *20,000 Leagues Under the Sea*, *Shopgirl*, *The Adventures of Huckleberry Finn*, and *Rogue One*. Pixel anomaly colors include black, blue, red, and white.

3.2 Model and Process

The NPED autoencoder (AE) is composed of single dense encoder and decoder layers, with an embedding dimension of 32, RELu activation function, and a binary crossentropy loss function. Reconstruction error (RE) is defined as the Euclidean distance between the original and reconstructed patches.

For testing, each video is broken into patches for each frame, fed to the autoencoder for reconstruction, and RE calculated. Local Outlier Factor (LoF) outliers are calculated from the distribution of all RE scores on the frame, and outliers scores compared to the preceding frame's corresponding patches to determine newness. New outliers are flagged as pixel errors.

3.2.1 CNN alternative. We also compare to a binary class CNN alternative trained on anomalous and non-anomalous pixels using artificially generated anomalous pixels. The CNN-temp model was trained on image patches with temporal differences between the current patch and its preceding counterpart encoded as an additional channel. For this method, frames were first fed through a low-pass filter and the highest interquartile range values from the filtered vs. original differences used to identify potential outliers. These outliers were then fed to the binary CNN for evaluation.

4 RESULTS

Results of the different AE patch models compared to the CNN model are presented in Table 1. Each test video contained one pixel anomaly, rendering recall a binary yes/no. False positive rates are presented at the frame level: the number of frames that contain false positives out of the total number of true non-anomalous frames.

The best model was the AE-10x10 pixel patch model, achieving 100% recall and an average false positive rate of 8.67%, with the highest false positive rate below a reasonable 20%.

5 SUMMARY

In sum, we show that simple neural network methods such as autoencoders can perform admirably for real-world production QC tasks, improving scalability for a studio QC workflow.

While the current NPED-AE 10x10 model performs acceptably on our test data, further testing on additional content is required. In addition, workflow improvements may be achieved by incorporating object and facial detection layers to only flag potential anomalies that occur on areas of high salience for the audience, such as character faces.

Further investigations would also be fruitful in other modeling techniques or combinations, such as combining the autoencoder with binary classification, investigating inpainting approaches, and expanding scope to animation as well as live-action video.

REFERENCES

Jinwon An and Sungzoon Cho. 2015. Variational Autoencoder based Anomaly Detection using Reconstruction Probability.

Longin Jan Latecki, Aleksandar Lazarevic, and Dragoljub Pokrajac. 2007. Outlier Detection with Kernel Density Functions. In *Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM '07)*. Springer-Verlag, Berlin, Heidelberg, 61–75. https://doi.org/10.1007/978-3-540-73499-4_6

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1. MIT Press, Cambridge, MA, USA, Chapter Learning Internal Representations by Error Propagation, 318–362. http://dl.acm.org/citation.cfm?id=104279.104293

Haowen Xu, Wenxiao Chen, Nengwen Zhao, Zeyan Li, Jiahao Bu, Zhihan Li, Ying Liu, Youjian Zhao, Dan Pei, Yang Feng, Jie Chen, Zhaogang Wang, and Honglin Qiao. 2018. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 187–196. https://doi.org/10.1145/3178876.3185996