Can a CNN Detect Professional Image Forgery?

Achilleas Vlogiaris Arkajit Bhattacharya Kyriakos Psarakis Panagiotis Soilis Rafail Skoulos
Delft University of Technology

1. Research Topic

- Image forgery detection is becoming more and more challenging (Social media + manipulation software)
- The CNN model proposed by [1] achieves an impressive accuracy of more than 98% on the CASIA dataset
- However, CASIA contains images that are relatively easy to recognize by humans (Figure 1)

Figure 1: CASIA example

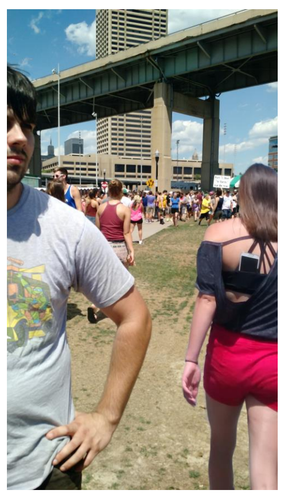

Figure 2:
Difficult dataset[2] example

Research questions:

1. How does the dataset difficulty influence classification performance? Are more challenging datasets for humans also more demanding for a CNN?
2. How do different optimization hyperparameters influence the model output?

2. Technical Approach

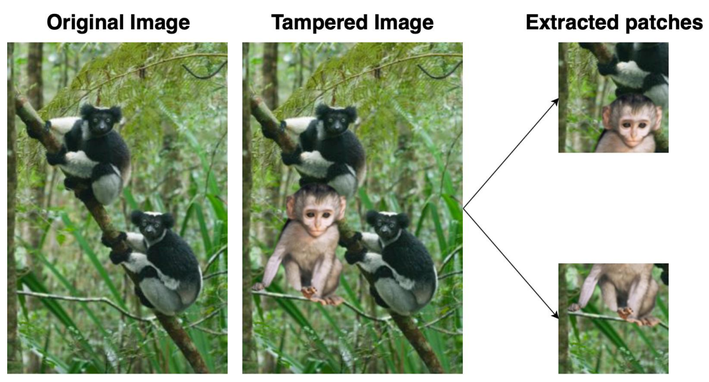

Figure 3: Patch extraction example

Patch extraction:

- Extract patches 128x128px from tampered/authentic images in CASIA2 and NC16 [2] datasets
- For each tampered image extract two random patches from the forged region
- For each authentic image extract two random patches

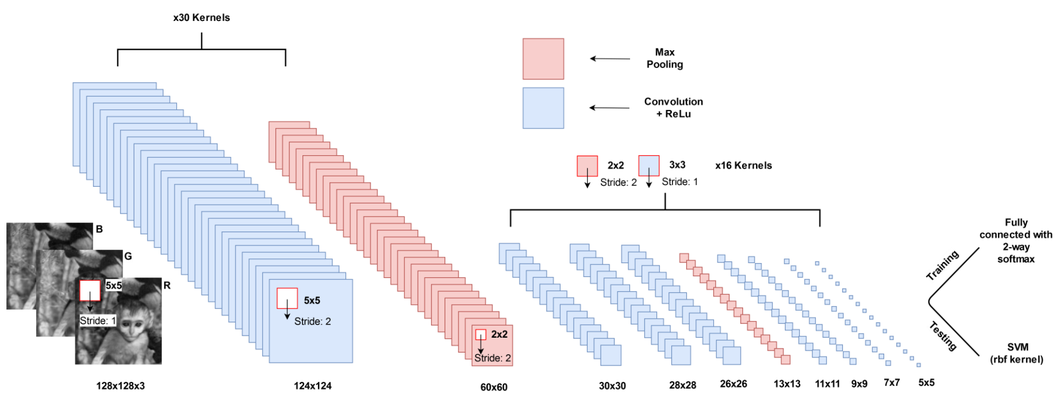

Figure 4: CNN architecture overview

CNN architecture:

- 10 layers: 8 convolutions + 2 max pooling
- Training: input patches, fully connected layer, softmax
- Testing: input patches, SVM

3. Experimental Setup

- # patches extracted without data augmentation:
  - CASIA2: 20,076 images
  - NC16: 2,204 images
- First convolutional layer: SRM initialization filters
- CNN hyperparameters optimized per dataset
- Mean feature fusion used
- Classification via SVM with RBF kernel
- SVM accuracy error via 10-fold cross-validation

4. Results

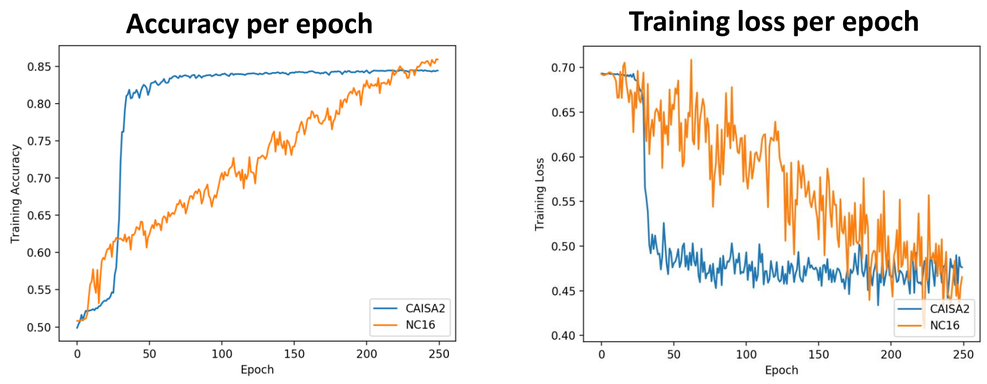

Figure 5: CASIA2 vs NC16 - CNN training behavior

CNN SGD hyperparameters: Momentum=0.99, Weight Decay=0.0005
CASIA2: No dropout, batch=200, LR: 0.0005 vs NC16: 50% dropout, batch=32, LR: 0.001

10-fold cross-val	Accuracy	Comparison	Std
CASIA2	92.54%	-	3.81%
NC16	83.29%	-10%	5.59%

Table 1: Dataset accuracy comparison

SVM hyperparameters: CASIA2: gamma= 0.0001, C=1 vs NC16: gamma= 0.001, C=100

CASIA2	Predicted Authentic	Predicted Tampered
Actual Authentic	1,308	190
Actual Tampered	12	1,013

NC16	Predicted Authentic	Predicted Tampered
Actual Authentic	88	24
Actual Tampered	13	100

Table 2: Confusion matrices with 80-20 split

5. Discussion

Research findings:

1. The more difficult NC16 dataset degrades the CNN performance by around 10%* compared to CASIA2.
2. Wrong hyperparameters = CNN cannot learn
3. There are more FP than FN

* The NC16 contains 1/10 of the CASIA2 training patches

Learnings:

- Deploying on Google Cloud can be challenging
- Scaling a model to more data is non-trivial
- CNN training is very sensitive to hyperparameters
- Academic papers can contain contradicting information

Remaining experiments:

- Test on CASIA1: similar size to NC16
- Use data augmentation to “obtain” more data
- Experiment with dropout and early stopping
- Further optimize hyperparameters

References

[1] Yuan Rao and Jiangqun Ni. A deep learning approach to detection of splicing and copy-move forgeries in images. In 2016 IEEE International Workshop on Information Forensics and Security(WIFS), pages 1–6. IEEE, 2016.
[2] Haiying Guan, Mark Kozak, Eric Robertson, Yooy-oung Lee, Amy N Yates, Andrew Delgado, DanielZhou, Timothee Kheyrkhah, Jeff Smith, and JonathanFiscus. Mfc datasets: Large-scale benchmark datasetsfor media forensic challenge evaluation. In 2019 IEEE Winter Applications of Computer Vision Workshops(WACVW), pages 63–72. IEEE, 2019.

TU Delft