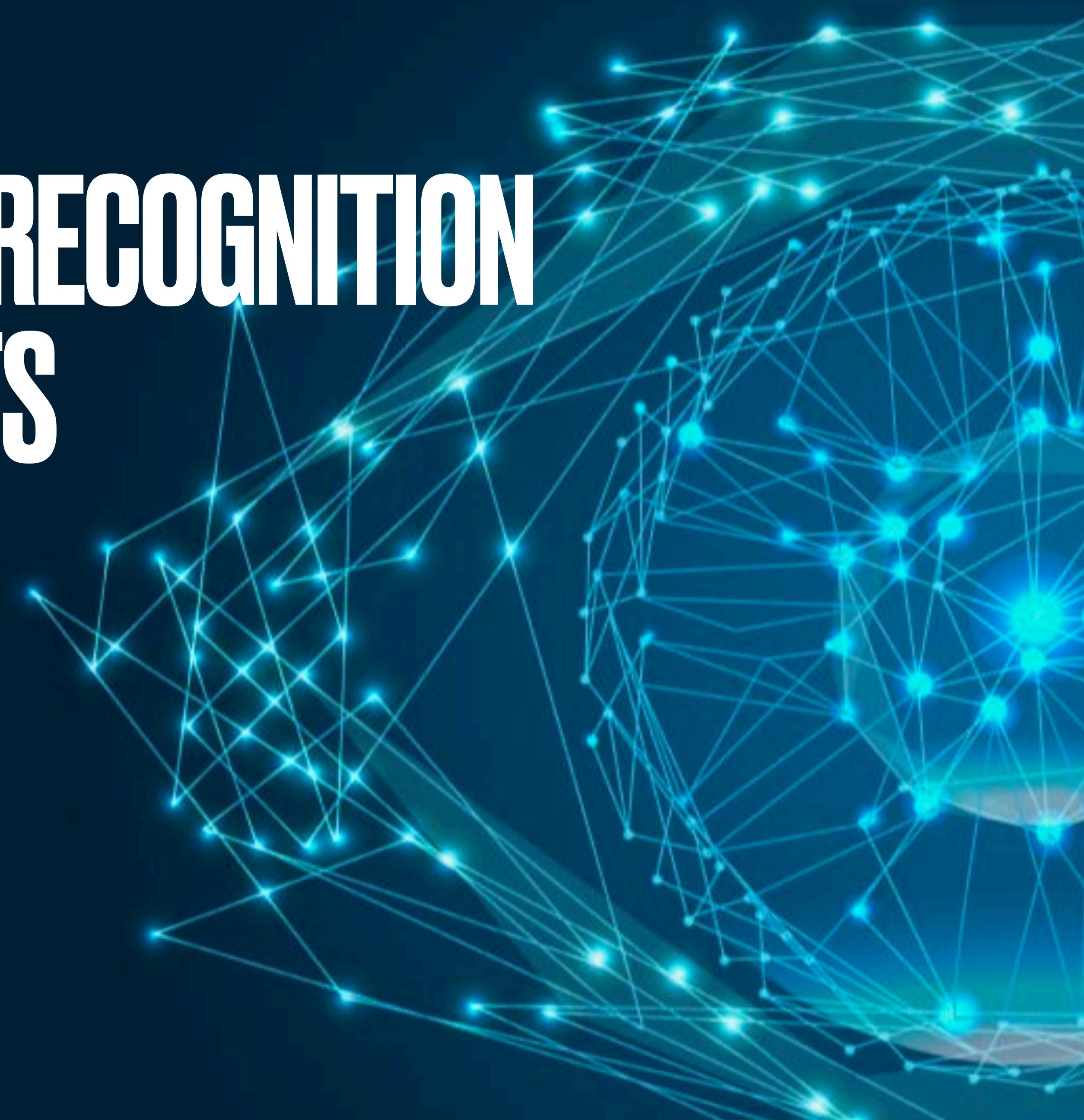


WALTER TYRNA | METIS

DEEP LEARNING FOR TEXT RECOGNITION IN ILLICIT ADVERTISEMENTS

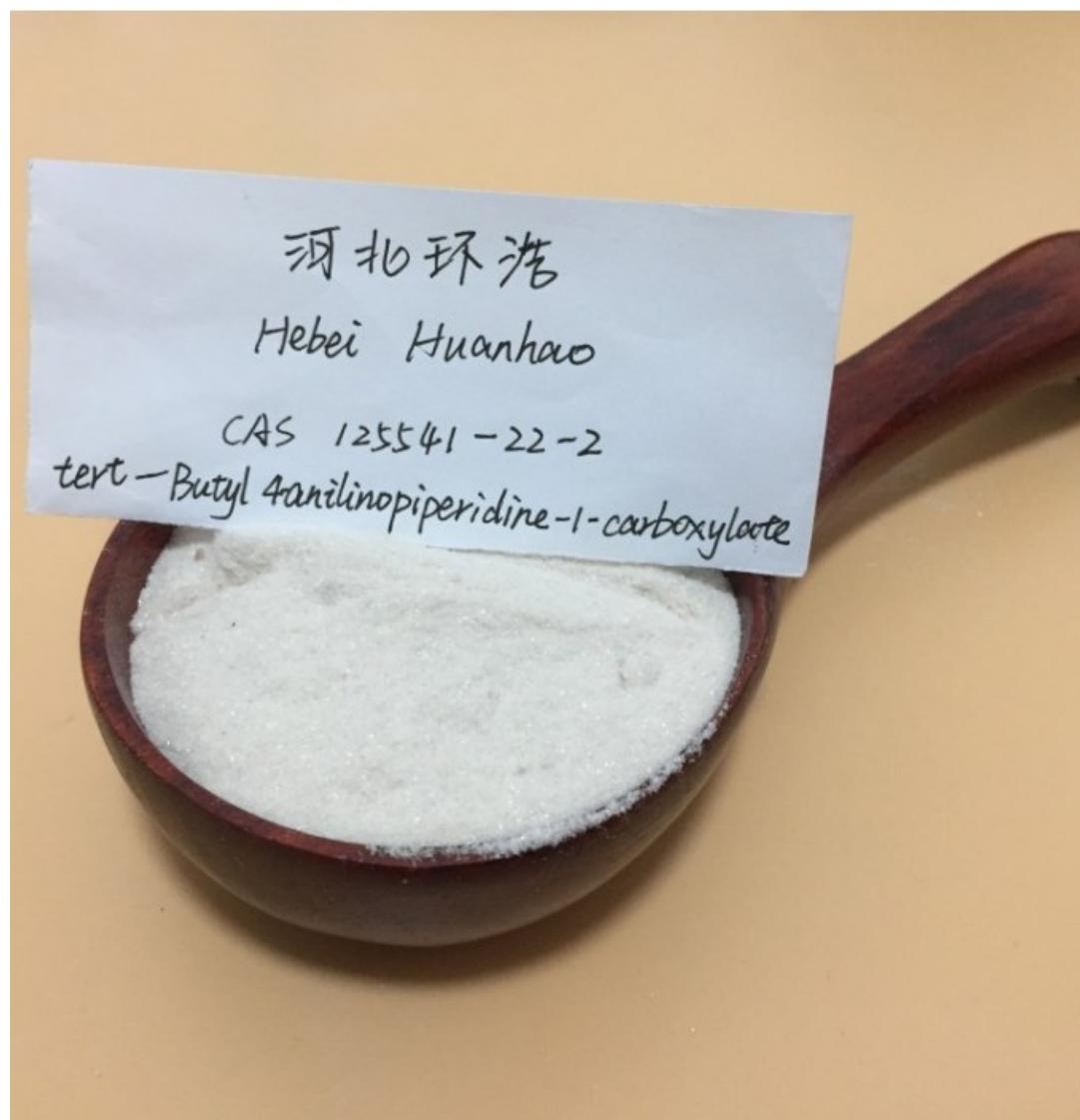


OVERVIEW

- Problem Introduction & Proposed Solution
- Data
- Methodology
- Results
- Future Work
- Appendix

EMBEDDED TEXT IN ILLICIT CONTENT

- Problem: Illicit actors often embed text in advertisements to avoid web scrapes, adding hours to data collection and analysis.



- Example: Chemical suppliers provide contact and product information in advertisements for synthetic drug precursors

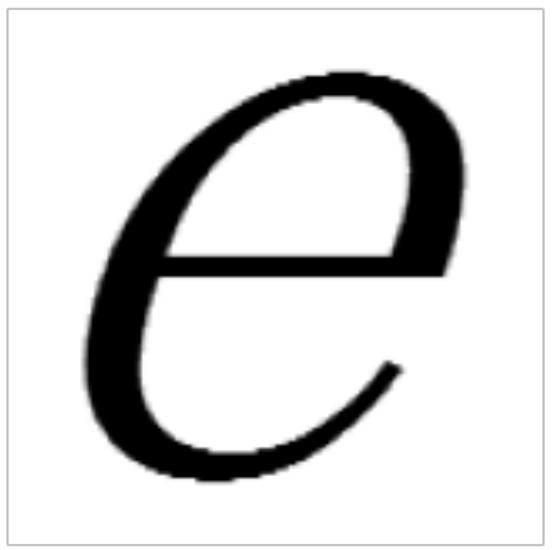
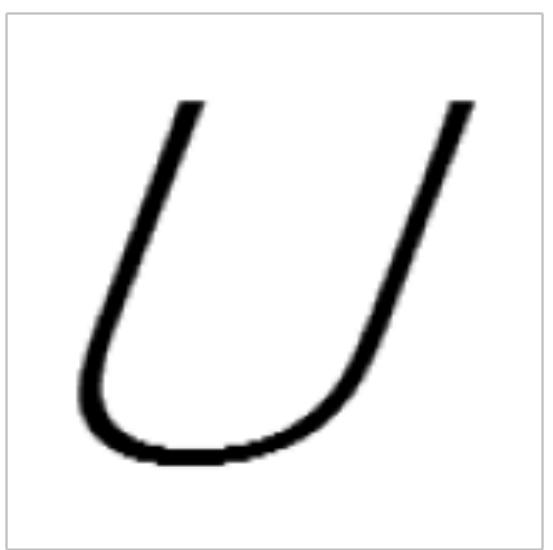
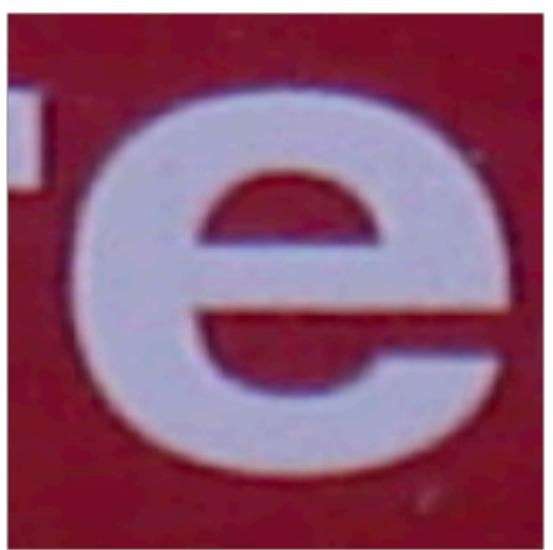
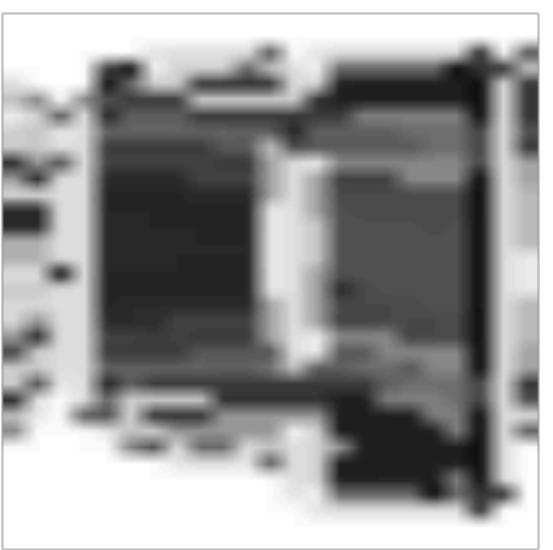
EMBEDDED TEXT IN ILLICIT CONTENT

- Proposed Solution: Use a deep learning approach to extract and identify text in advertisements
- Current Project Scope: Create a model that can identify individual characters from real life images

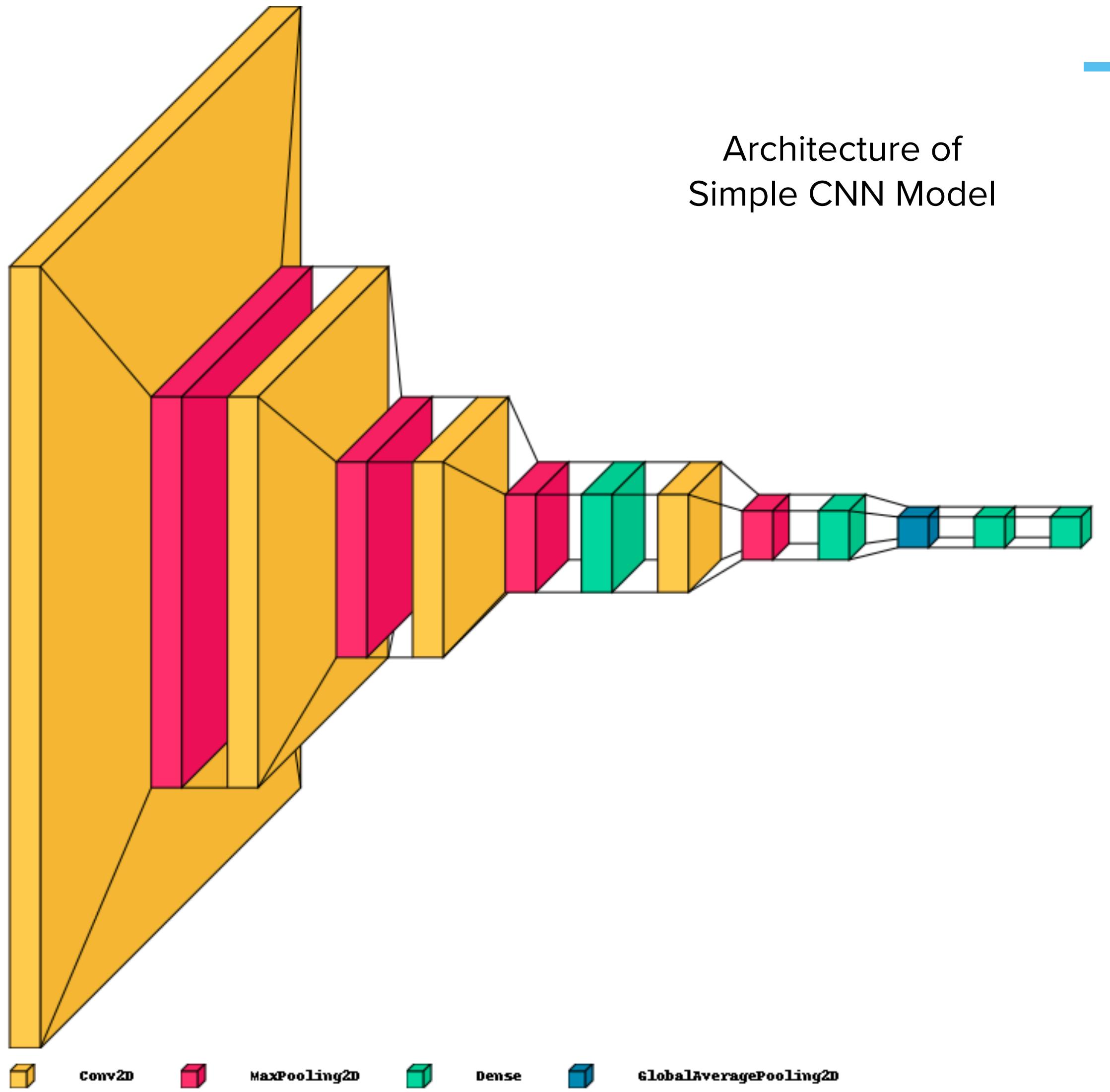


DATA

- Chars 74k Dataset:
 - 62 classes (0-9, A-Z, a-z)
 - 74,185 images:
 - 7705 characters from real life images
 - 3410 hand drawn characters
 - 62992 characters from computer fonts

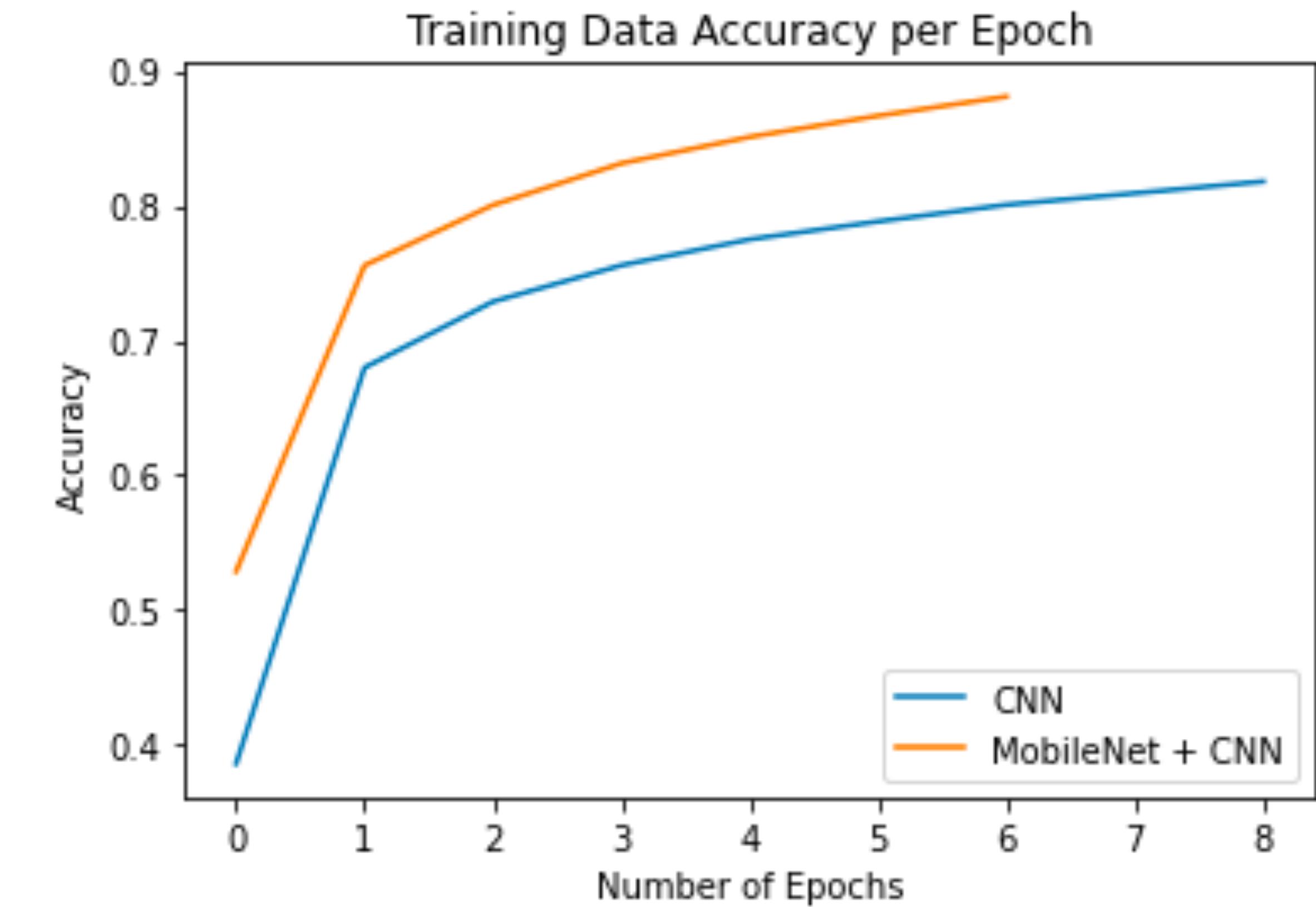
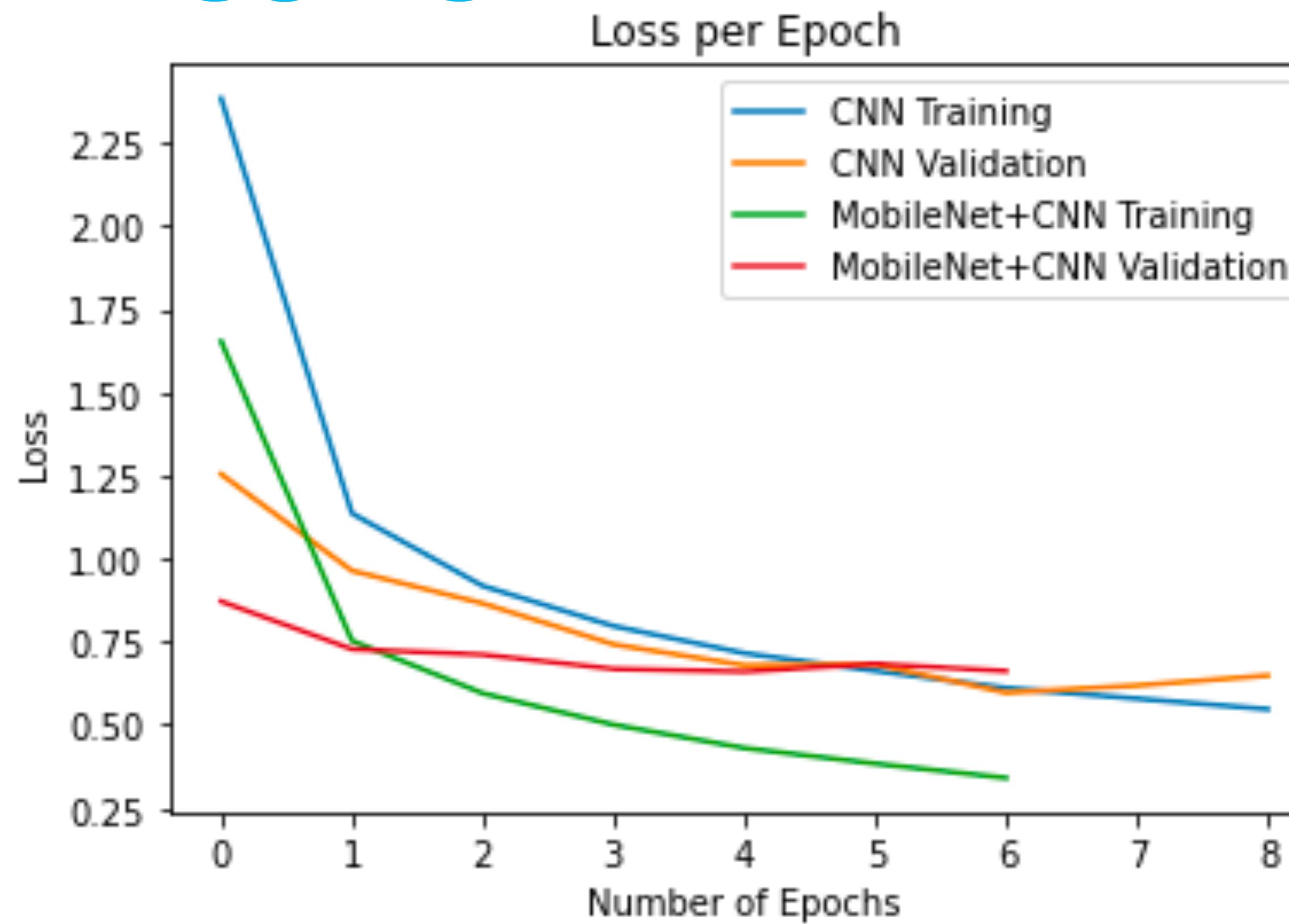


METHODOLOGY



- Two Neural Networks:
 - Simple Convolutional Neural Network (CNN)
 - Transfer Learning: Mobile Net + Simple CNN

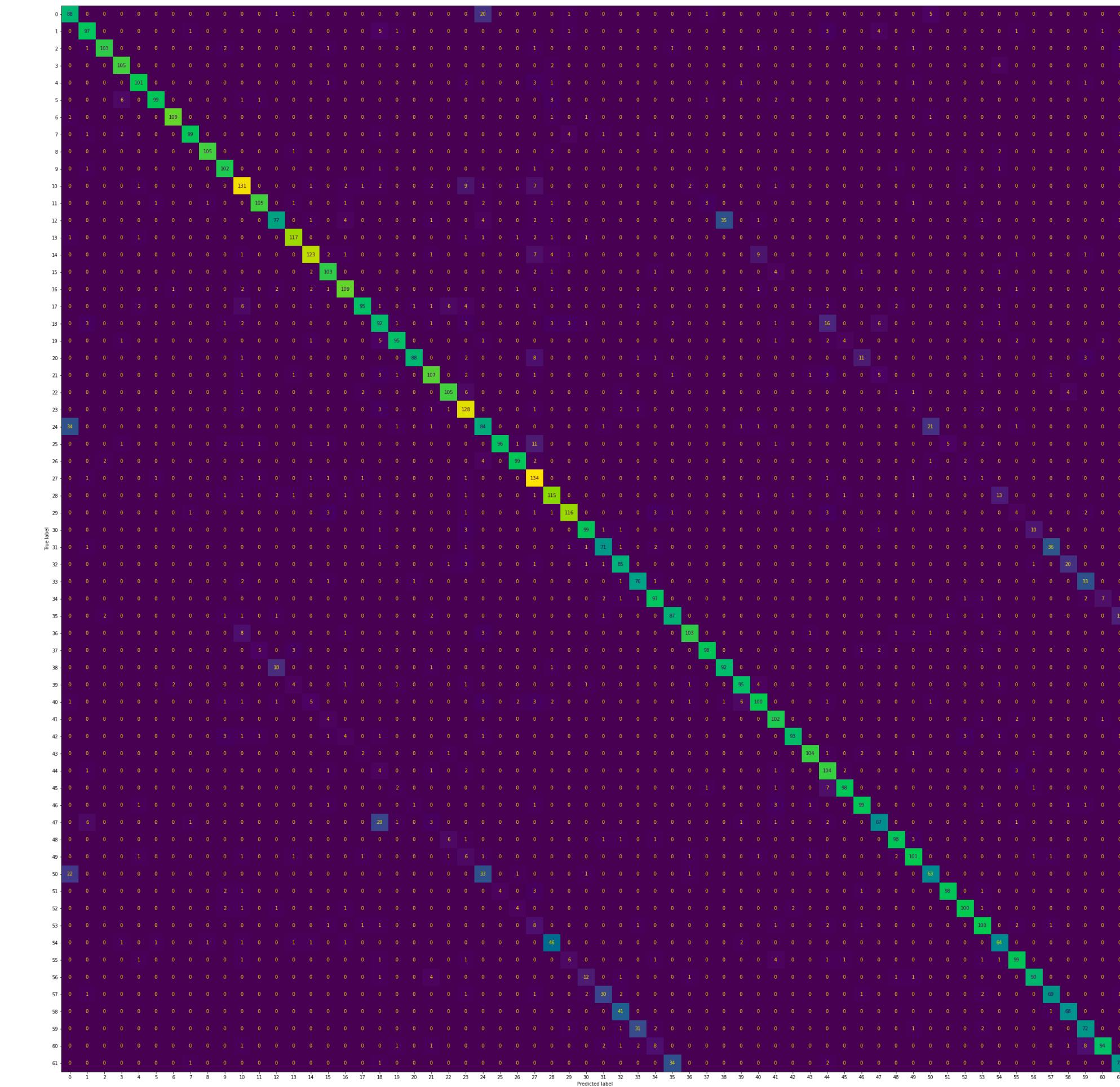
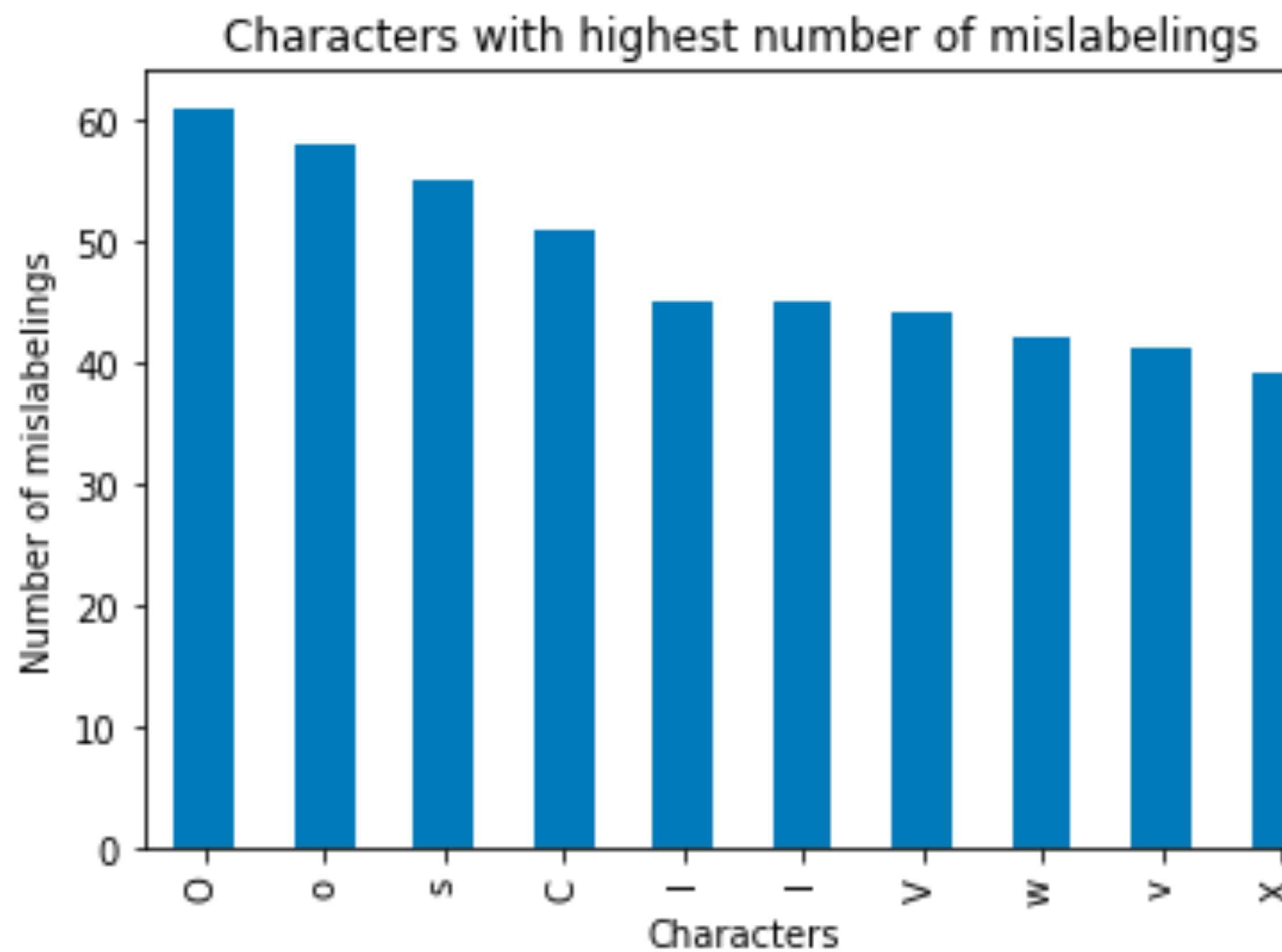
RESULTS



Simple CNN Training Accuracy: 0.818

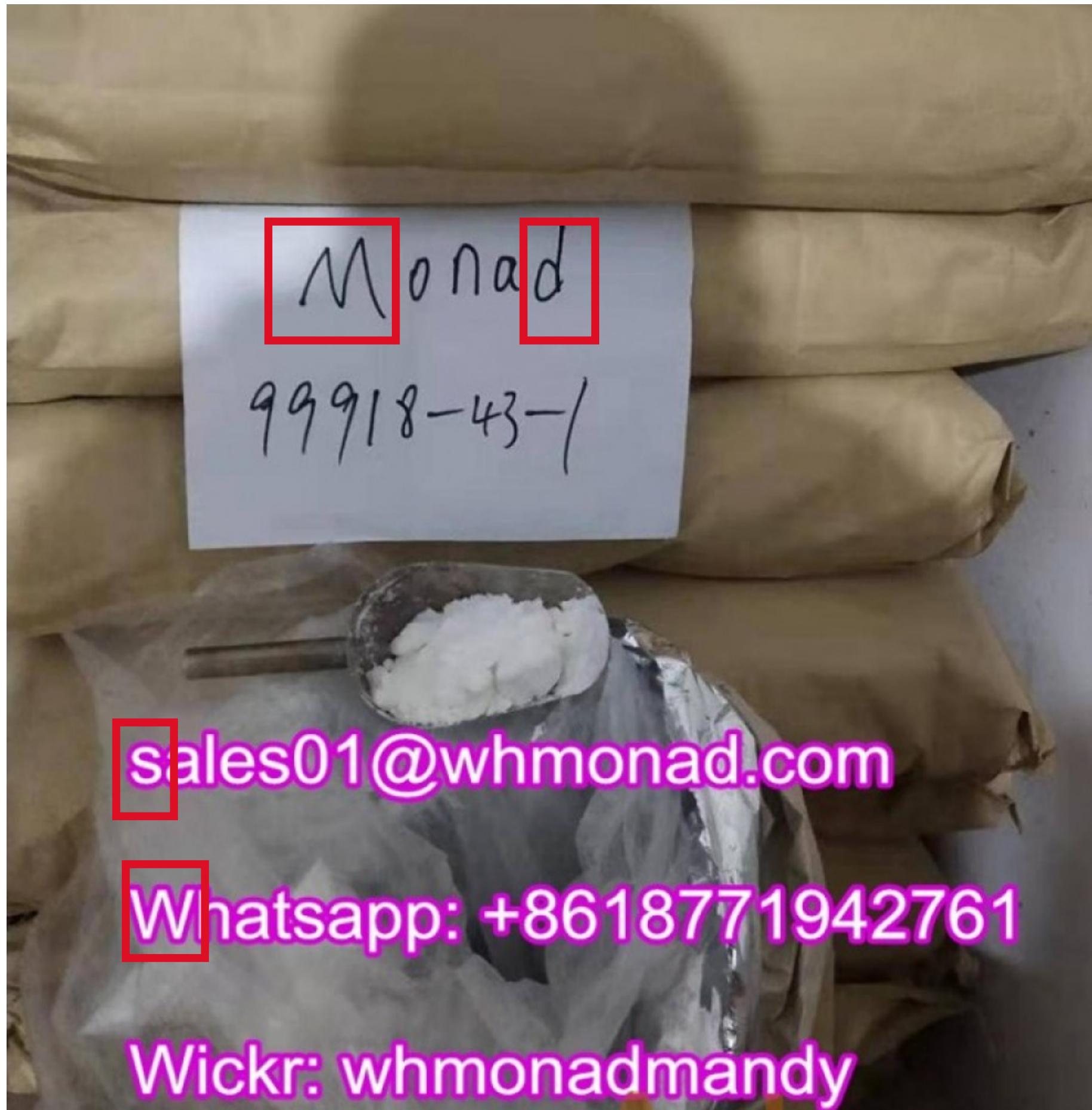
Mobile Net + SCNN Training Accuracy: 0.878

RESULTS



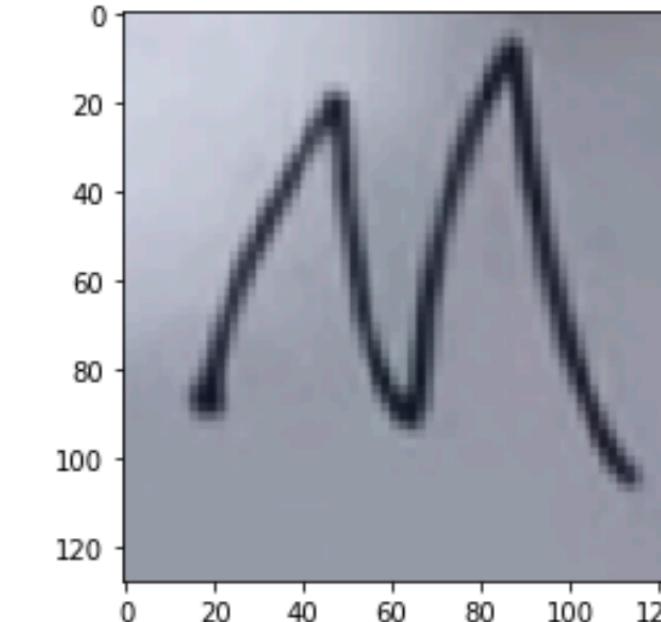
[Link to interactive confusion matrix](#)

RESULTS



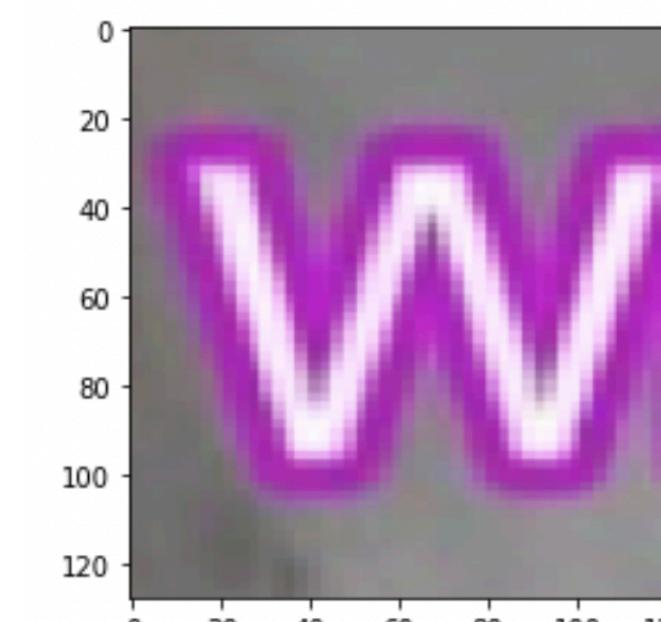
```
real_life_images('Real_Life_Text/monad_m.png')
```

0 The model predicts the image to be: m
dtype: object



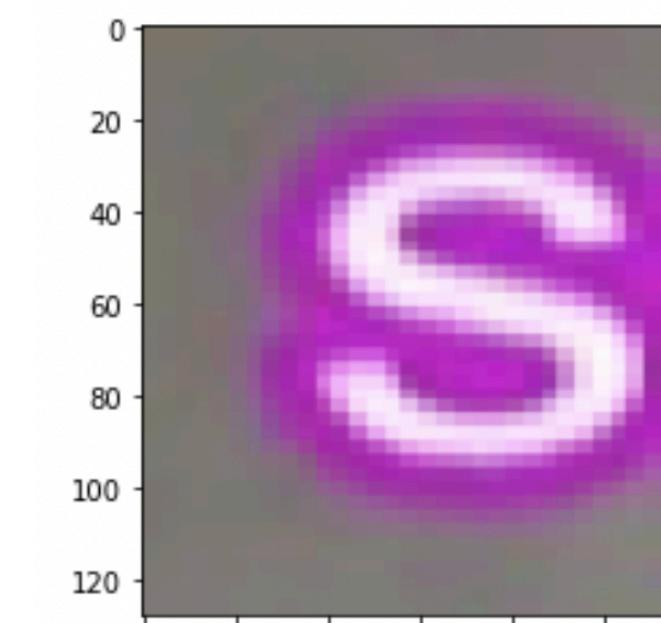
```
real_life_images('Real_Life_Text/monad_w.png')
```

0 The model predicts the image to be: 4
dtype: object



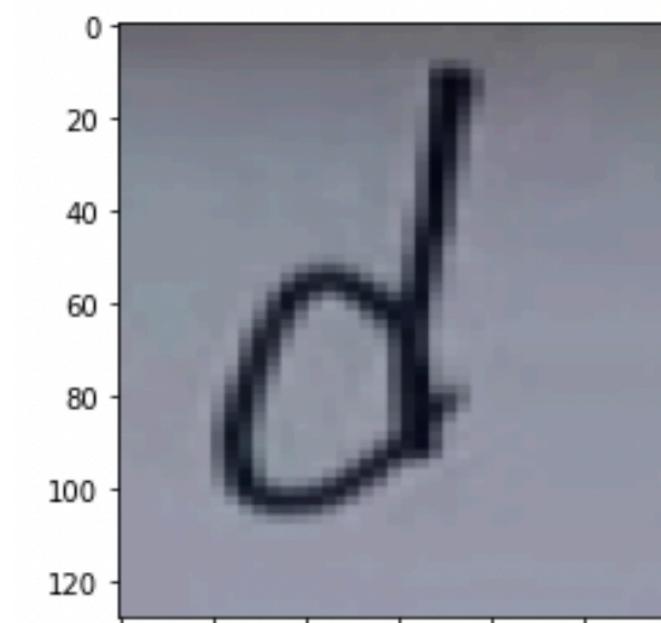
```
real_life_images('Real_Life_Text/monad_s.png')
```

0 The model predicts the image to be: S
dtype: object



```
real_life_images('Real_Life_Text/monad_d.png')
```

0 The model predicts the image to be: d
dtype: object



FUTURE WORK

- Better differentiation between similar characters
- Incorporation of OpenCV
- Scrape & addition of domain-specific text images

APPENDIX

