



AUTOMATING CONTENT FILTERING

Carl Cook

The Problem



Convolutional Neural Networks

Input
(images)

Output
(predictions)



Feature
Detection

Convolutional Neural Networks



input

convolutions

output



Confusion matrix

Actual	E	1293 14.41%	25 0.28%	108 1.20%	824 9.19%	2250 57.47% 42.53%
	E10plus	903 10.07%	20 0.22%	80 0.89%	897 10.00%	1900 1.05% 98.95%
	M	344 3.84%	5 0.06%	490 5.46%	971 10.82%	1810 27.07% 72.93%
	T	968 10.79%	15 0.17%	269 3.00%	1758 19.60%	3010 58.41% 41.59%
	sum_col	3508 36.86% 63.14%	65 30.77% 69.23%	947 51.74% 48.26%	4450 39.51% 60.49%	8970 39.70% 60.30%
		Predicted				
		E	E10plus	M	T	sum_lin

Simple Neural Network

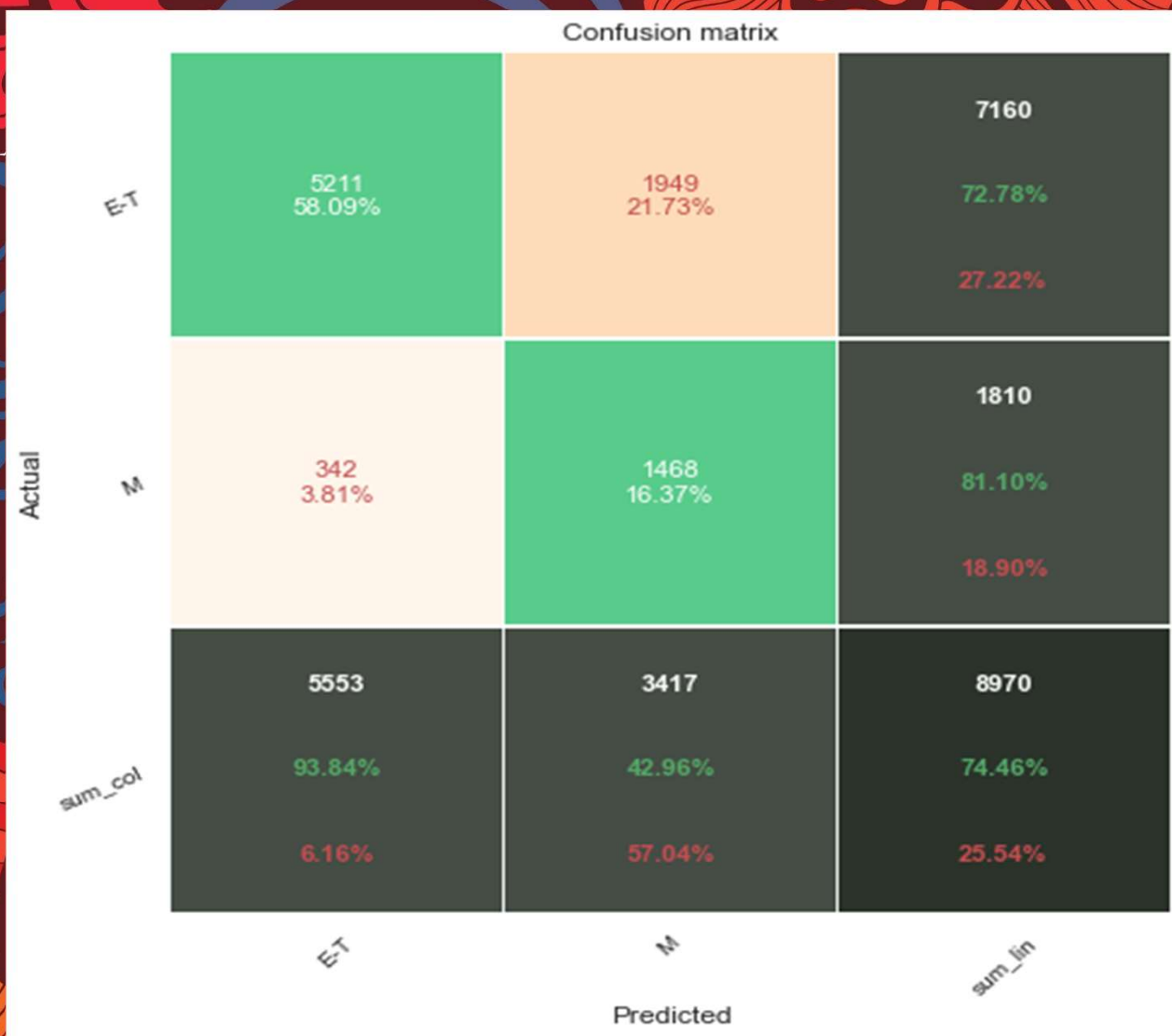
- 40% accuracy
- 27% recall
- 52% precision

Confusion matrix

Actual	E	1280 14.27%	224 2.50%	202 2.25%	544 6.06%	2250 56.89% 43.11%
	E10plus	658 7.34%	277 3.09%	138 1.54%	827 9.22%	1900 14.58% 85.42%
	M	211 2.35%	25 0.28%	1076 12.00%	498 5.55%	1810 59.45% 40.55%
	T	620 6.91%	165 1.84%	640 7.13%	1585 17.67%	3010 52.66% 47.34%
	sum_col	2769 46.23% 53.77%	691 40.09% 59.91%	2056 52.33% 47.67%	3454 45.89% 54.11%	8970 47.02% 52.98%
		Predicted				
		E	E10plus	M	T	sum_lin

Convolutional Neural Network

- 47% accuracy
- 59% recall
- 52% precision



VGG-16 Binary

- 74% accuracy
- 81% recall
- 43% precision

Conclusion

A step toward automated content filtering

Useful for automated content flagging

Human review still needed

Recommendations

Consistent gameplay images

Predictions based on multiple images

Video clip classification



Future Work

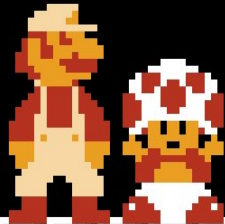
NLP for monitoring chat

Monitoring audio

Video classification

THANK YOU MARIO!

BUT OUR PRINCESS IS IN
ANOTHER CASTLE!



Thank You!

Your time and attention are
appreciated!