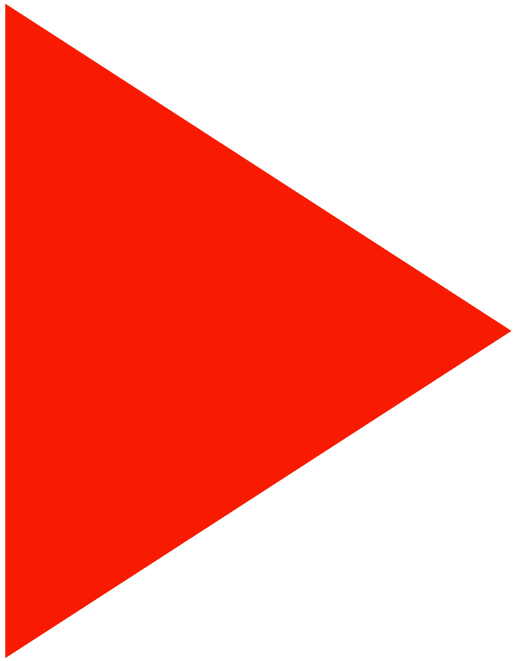


# YouTube

Trending Videos Categorization with Deep Learning

# Can machine categorize videos simply base on their thumbnails and video titles?

- Question/hypotheses

# The task:

# How hard is it?

Thumbnail:


+

Title:

Founding An Inbreeding-Free Space Colony

- Friend 1:
- "Science and Technology"

- Friend 2:
- "Gaming"

- Grand Truth:
- "Education"

Categories:

1. Entertainment
2. Music
3. How to and Style
4. Comedy
5. News and Politics
6. People and Blogs
7. Sport
8. Science and Technology
9. Film and Animation
10. Education
11. Pets and Animals
12. Gaming
13. Auto & Vehicles
14. Travel and Events

# Human-level Performance

## Experiment results with my friend (~100 trials):

WITH ONLY IMAGES:

30% ACCURACY

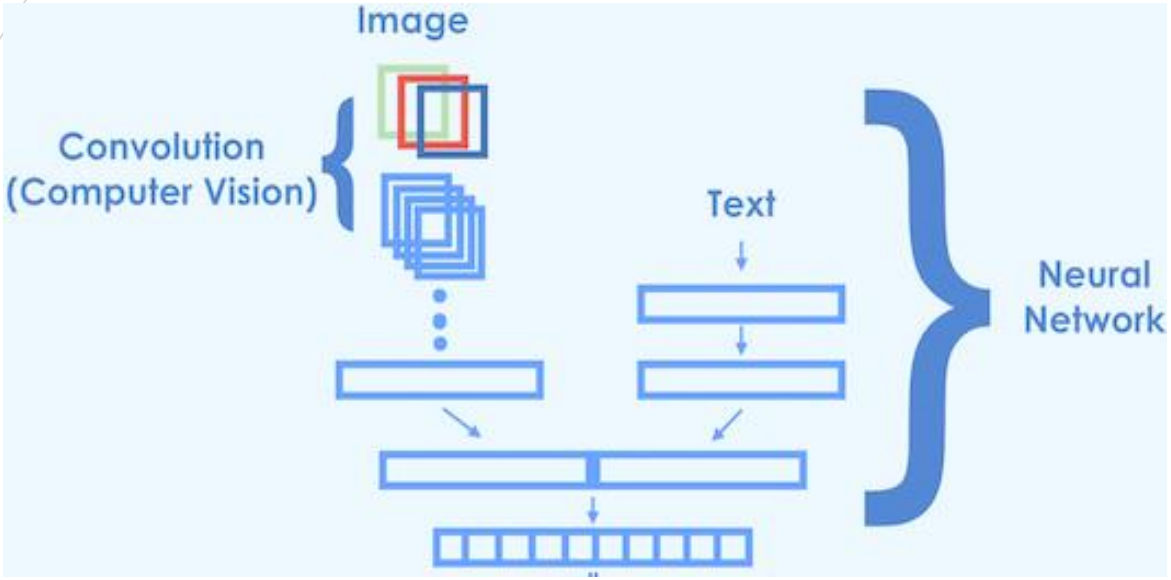
WITH ONLY TITLES:

50% ACCURACY

WITH BOTH IMAGES AND TITLES:

54% ACCURACY

## NLP (Titles) + CV (Thumbnails)




1. Build separate models as a independent NLP/CV problem

2. Then combining two models with one dense output layer

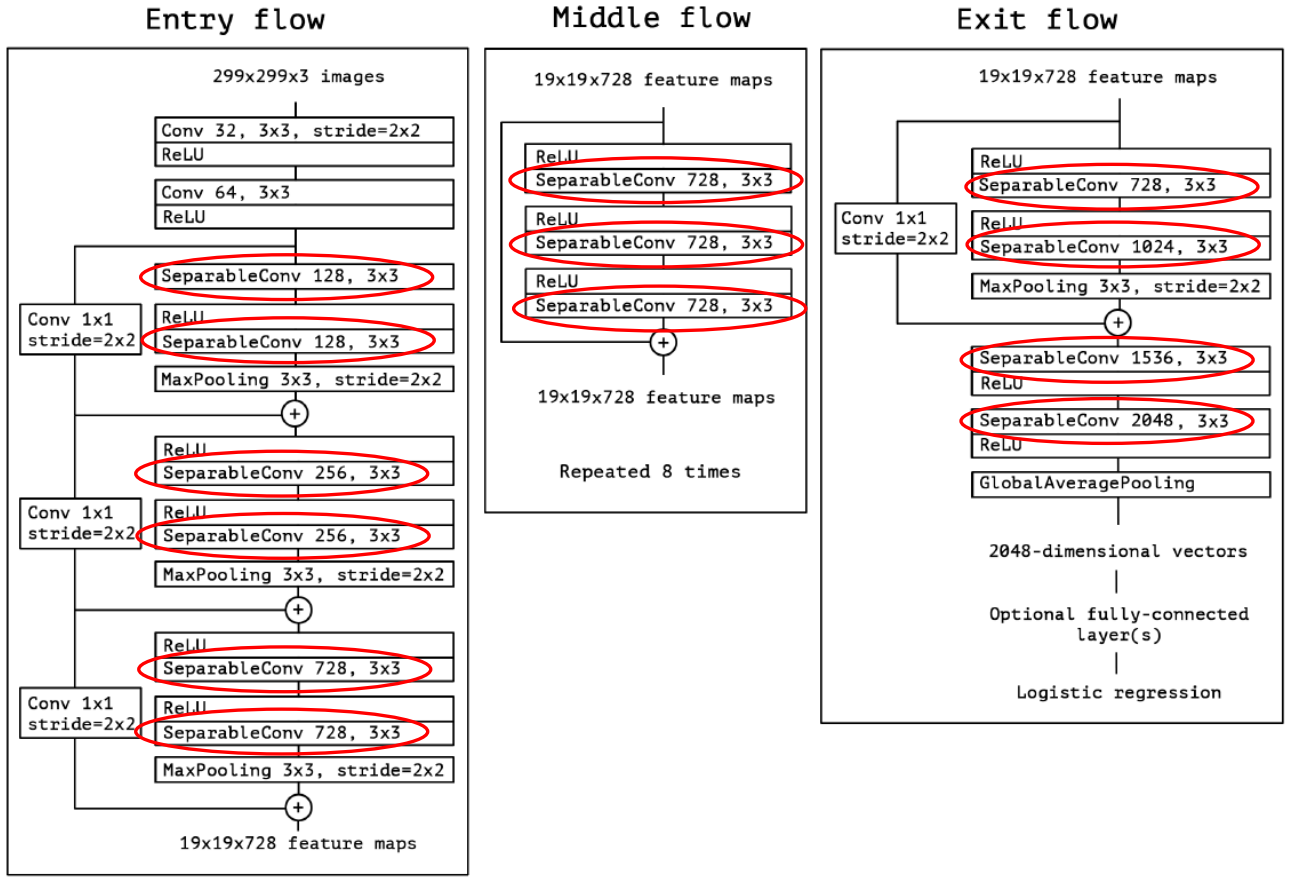
# Data (5,000 training data + 1,020 test data)

Labels - 14 unique video categories, converted into one hot vector format.

Image data - 120 x 90 x 3 matrix with the numbers representing the brightness of the 120 x 90 pixies for each of the RGB channels.

Text data – Video titles, each word in the title will be converted to a 50-dimensional vectors by using word embedding pre-trained on 2 billion tweets with 1.2 million vocabularies.

# Approach for Image model
## Xception model with transfer learning

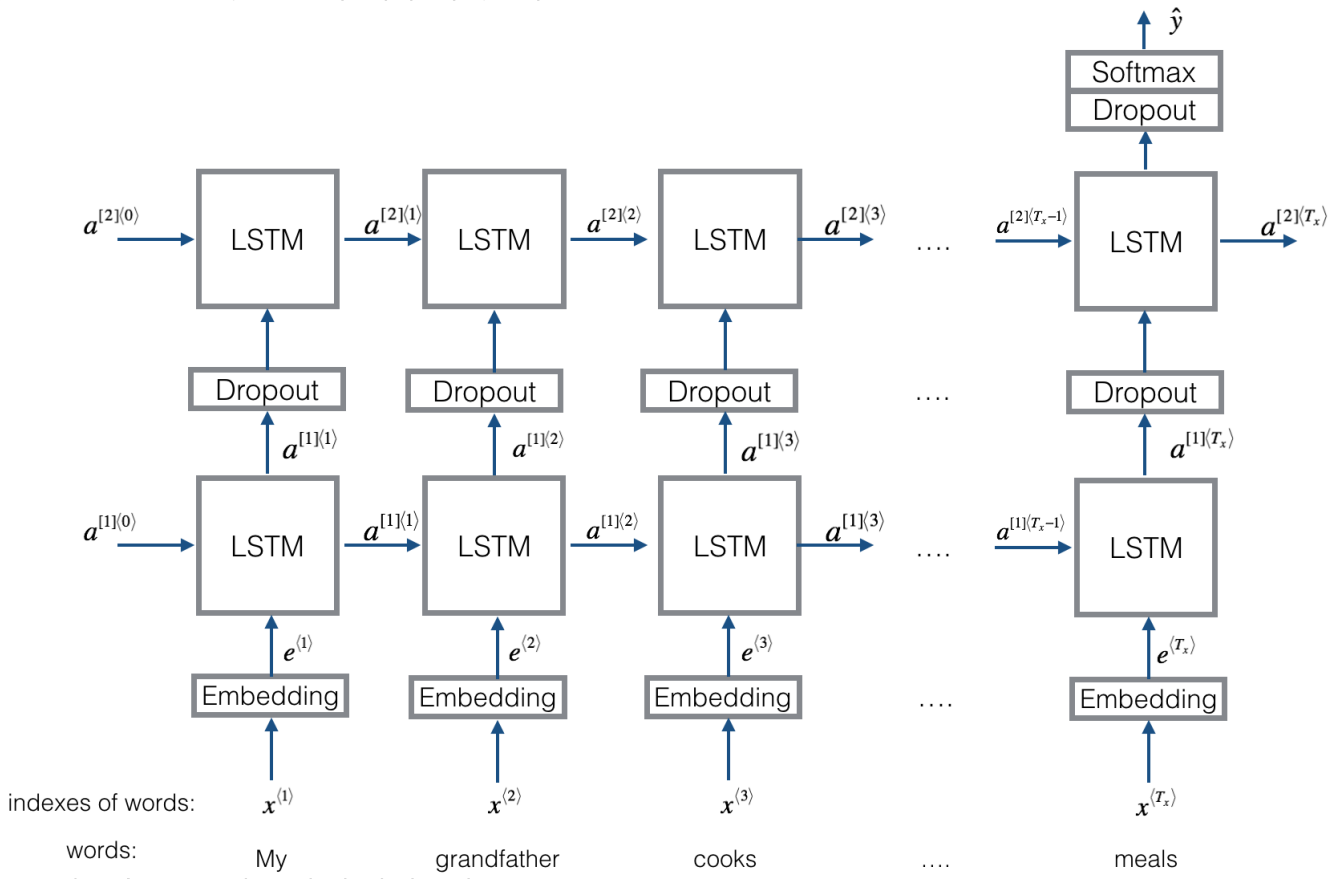



| Layer (type) | Output Shape | Param # |
|---|---|---|
| xception (Model) | (None, 3, 3, 2048) | 20861480 |
| flatten_9 (Flatten) | (None, 18432) | 0 |
| dense_33 (Dense) | (None, 1024) | 18875392 |
| dropout_25 (Dropout) | (None, 1024) | 0 |
| dense_34 (Dense) | (None, 512) | 524800 |
| dropout_26 (Dropout) | (None, 512) | 0 |
| dense_35 (Dense) | (None, 128) | 65664 |
| dropout_27 (Dropout) | (None, 128) | 0 |
| dense_36 (Dense) | (None, 14) | 1806 |

Total params: 40,329,142
Trainable params: 19,522,190
Non-trainable params: 20,806,952

**Optimizer: Adam (adaptive moment)**
**Regularization: Dropout layers**

# Approach for Text model
# 2 layers LSTM sequence classifier




| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_7 (InputLayer) | (None, 20) | 0 |
| embedding_5 (Embedding) | (None, 20, 50) | 20000050 |
| lstm_7 (LSTM) | (None, 20, 128) | 91648 |
| dropout_12 (Dropout) | (None, 20, 128) | 0 |
| lstm_8 (LSTM) | (None, 128) | 131584 |
| dropout_13 (Dropout) | (None, 128) | 0 |
| dense_11 (Dense) | (None, 14) | 1806 |
| activation_4 (Activation) | (None, 14) | 0 |

Total params: 20,225,088
Trainable params: 225,038
Non-trainable params: 20,000,050

**Optimizer: Adam (adaptive moment)**
**Regularization: Dropout layers**

# Goals

Human-level performance

# Evaluation Metrics

Top-1 accuracy

Top-5 accuracy

Confusion matrix

# Top-1 & top-5 accuracy

## After fine tuning the hyperparameters:

WITH ONLY IMAGES:

TOP-1: 30.39% ACCURACY

TOP-5: 73.14% ACCURACY

GOAL: 30%

WITH ONLY TITLES:

TOP-1: 51.18% ACCURACY

TOP-5: 85.78% ACCURACY

GOAL : 50%

WITH BOTH IMAGES AND TITLES:

TOP-1: 45.30% ACCURACY

TOP-5: 84.22% ACCURACY

GOAL: 54%



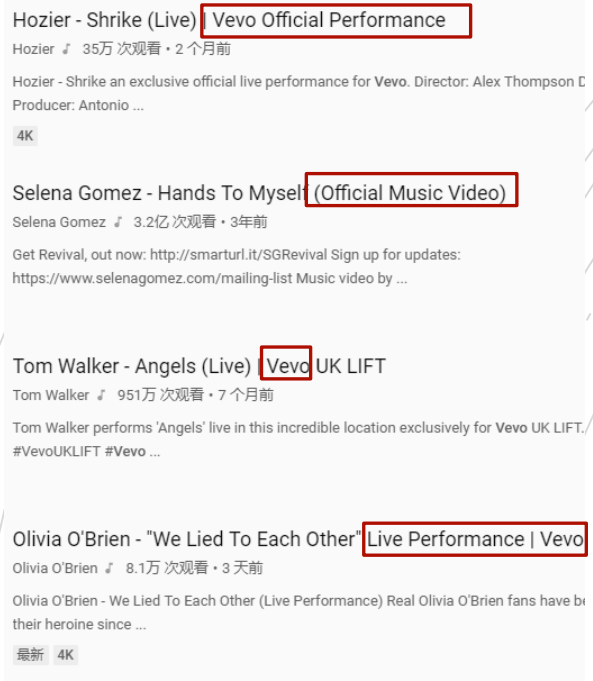


# 1. Confusion Matrix for Image model

Grand Truth

Film and Animation
Auto & Vehicles
Music
Pets and Animals
Sport
Travel and events
Gaming
People and Blogs
Comedy
Entertainment
News and Politics
How to and Style
Education
Science and Technology

```
[[ 0  0  5  0  2  0  0  1  4 36  0  3  0  2]
 [ 0  0  1  0  0  0  0  0  1  0  0  0  0  0]
 [ 1  0 45  0  1  0  0  2  3 82  1  3  0  0]
 [ 0  0  2  0  0  0  0  2  0 10  1  1  0  0]
 [ 0  0  6  0  0  0  0  3  4 54  2  4  0  0]
 [ 0  0  0  0  0  0  0  0  0  3  0  2  0  0]
 [ 0  0  2  0  0  0  0  1  1 26  1  6  0  1]
 [ 0  0 11  0  0  0  0  6  2 46  0 11  0  0]
 [ 0  0  1  1  0  0  0  2  5 73  0  6  0  1]
 [ 0  0 28  0  1  0  0  5  7 234  3 20  0  4]
 [ 0  0  6  0  1  0  0  1  1 31  1  4  0  2]
 [ 0  0  9  0  3  0  0  2  2 63  1 16  1  1]
 [ 0  0  4  0  3  0  0  2  1 13  0  9  1  1]
 [ 0  0  7  0  1  0  0  0  3 31  0  3  1  3]]
```




# 2. Confusion Matrix for Text model

Grand Truth

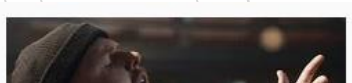
Film and Animation
Auto & Vehicles
Music
Pets and Animals
Sport
Travel and events
Gaming
People and Blogs
Comedy
Entertainment
News and Politics
How to and Style
Education
Science and Technology

```
[[ 26   0   3   1   1   0   2   2   2  12   1   0   0   3]
 [  0   0   0   0   0   1   0   0   0   0   0   0   0   1]
 [  2   1 100   0   1   0   0   3   4  21   1   0   1   4]
 [  1   0   0   6   0   1   1   1   2   1   0   1   1   1]
 [  0   0   1   0  62   2   1   0   1   4   1   0   0   1]
 [  0   1   0   0   0   1   0   0   0   0   0   2   0   1]
 [ 11   2   1   1   2   0   4   1   2   7   1   2   0   4]
 [  0   1   5   2   7   0   0   6   5  21   1  22   3   3]
 [  4   0   4   1   3   0   0  12  15  33   1   9   4   3]
 [ 21   0  11   1  12   0   1  21   4 181   7  27   6  10]
 [  0   0   1   1   1   2   0   1   2  12  17   0   4   6]
 [  1   0   1   0   2   1   0   9   1  23   0  52   1   7]
 [  0   1   1   0   0   0   0   1   2   4   0   3  10  12]
 [  0   1   2   2   1   0   0   2   3   5   0   0   3  30]]
```

# 3. Confusion Matrix for Combined model

Grand Truth

| | Films and Animation | Auto & Vehicles | Music | Pets and Animals | Sport | Travel and events | Gaming | People and Blogs | Comedy | Entertainment | News and Politics | How to and Style | Education | Science and Technology |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Film and Animation | 0 | 0 | 4 | 2 | 0 | 0 | 0 | 3 | 9 | 32 | 1 | 0 | 0 | 2 |
| Auto & Vehicles | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Music | 1 | 0 | 100 | 0 | 0 | 0 | 0 | 1 | 13 | 16 | 2 | 0 | 0 | 5 |
| Pets and Animals | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 15 |
| Sport | 0 | 0 | 0 | 0 | 56 | 0 | 0 | 2 | 3 | 3 | 4 | 0 | 0 | 5 |
| Travel and events | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 1 |
| Gaming | 0 | 0 | 1 | 1 | 2 | 0 | 1 | 2 | 6 | 13 | 1 | 0 | 0 | 11 |
| People and Blogs | 0 | 0 | 4 | 0 | 2 | 0 | 0 | 13 | 14 | 10 | 2 | 18 | 0 | 13 |
| Comedy | 0 | 0 | 3 | 0 | 2 | 0 | 0 | 17 | 29 | 21 | 2 | 6 | 0 | 9 |
| Entertainment | 0 | 0 | 16 | 0 | 7 | 0 | 0 | 20 | 51 | 145 | 3 | 20 | 0 | 40 |
| News and Politics | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 5 | 9 | 21 | 0 | 0 | 6 |
| How to and Style | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 19 | 6 | 8 | 0 | 47 | 0 | 17 |
| Education | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 2 | 3 | 0 | 4 | 0 | 21 |
| Science and Technology | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 3 | 3 | 0 | 2 | 0 | 36 |



Tom Walker - Angels (Live) | Vevo UK LIFT

Tom Walker ♪ 951万 次观看 • 7 个月前

# Possible Future Directions

- Auto tagging models for videos
- Better architectures to connect text and image data
- More data to evaluate the model and human-level performance
- Deal with class imbalance