ML Tips and Tricks

Suleman Kazi

What this presentation is and isn’t

When and where to (and not to) use machine learning.

Tips for better, cheaper, datasets and models.

Interpretability.

No calculus, linear algebra, or optimization.

No API descriptions or code.

When and where to (and not to) use machine learning.

Tips for better, cheaper, datasets and models.

Interpretability.

No calculus, linear algebra, or optimization.

No API descriptions or code.

Assumptions:

- Some basic knowledge of using or applying deep learning.
- Assumes no math or graduate school level background.

Prosthetic Knowledge

- Information that a person does not know, but can access as needed using technology. Definition from twitter.

- This presentation has **references** not details**.** You have to go and look them up as/when needed.

Do you need Machine Learning?

Don't do it because everyone is

- Can you get your desired performance faster and cheaper with 'traditional' algorithms?
- Can you get large amounts of data?
- Are you ok with less interpretability?

Robot control: You can get millimeter level control without ML using just accurate mass, inertia measurements and simple matrix multiplies

Do you need just Machine Learning?

Hybrid Approaches

- ML+Traditional Algorithm Hybrids. (We'll talk about ML-ML hybrids as well later).
- Especially useful when you have a small amount of data.
- Helpful in constrained compute situations where ML-only might be too computationally expensive.

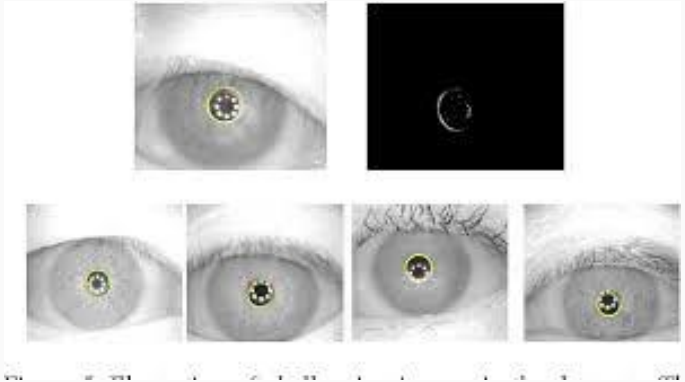
Examples

[Improving RANSAC-Based Segmentation Through CNN Encapsulation]

$$L = \log\left(\frac{1+S'}{1+S^*}\right) - \alpha \sum_{\substack{(x,y)\in C^* \\ Z(x,y)\leq 0}} Z(x,y) + \beta \sum_{\substack{(x,y)\notin C^* \cup C' \\ Z(x,y)>0}} Z(x,y) \quad (4)$$

With S' and S^* the scores of the strongest impostor C' and the true model C^* respectively, with scores computed by the following:

$$S = \sum_{\substack{(x,y)\in C \\ Z(x,y)>0}} Z(x,y) \quad (5)$$



Examples

Sentiment Analysis using Parts of Speech Tagging (ML Yearning, Andrew Ng.)

Preprocess text using PoS tagging before passing on to classifier:

This is a great movie -> This is a great $_{\text{Adjective}}$ movie $_{\text{Noun}}$!

Non-deep learning approaches

- Try to see how well some non deep-learning approaches work on your problem.
- Examples include SVM, Decision Trees, or maybe just plain old least squares.
- I have made performant hand tracking using just least squares. See this great course for more info: (EE263 Linear Dynamic Systems)

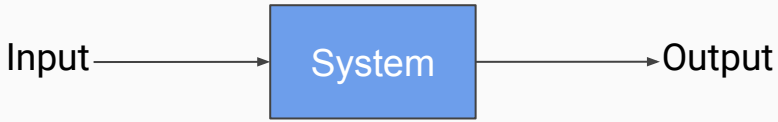
A Model Is Born

Divide and Conquer (Sometimes)

Break the problem down into parts.

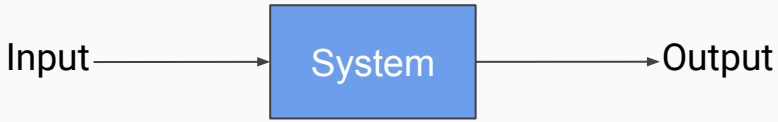
Divide and Conquer (Sometimes)

Break the problem down into parts.



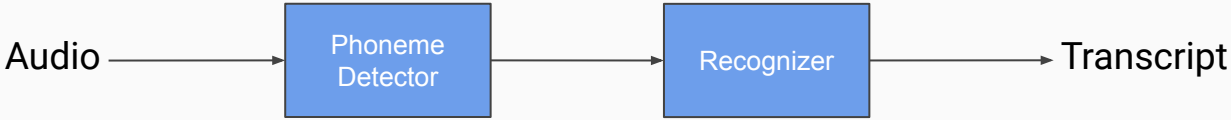
Divide and Conquer (Sometimes)

Break the problem down into parts.



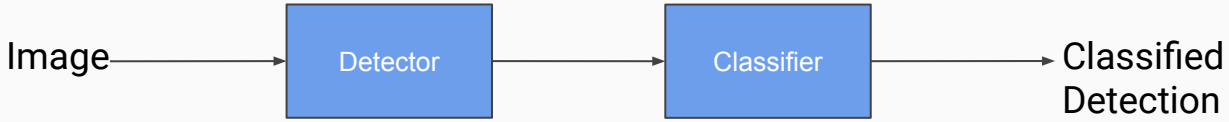
Divide and Conquer (Sometimes)

ASR (Automatic Speech Recognition)



Divide and Conquer (Sometimes)

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks



Build and fail fast (Initially)

- Start with a quick and dirty system, don't worry about the accuracy.
- This helps you set up a pipeline, see if there are any obvious errors and so on.
- Overfit on a tiny training set to see if your pipeline is working.

Visualize Every Step of the way

- Deep learning is annoyingly good, it will give you SOMEthing even in cases where your data pipeline is bad.

Visualize Every Step of the way

The dumb reason your fancy Computer Vision app isn't working: Exif Orientation

[Link to story]

Camera

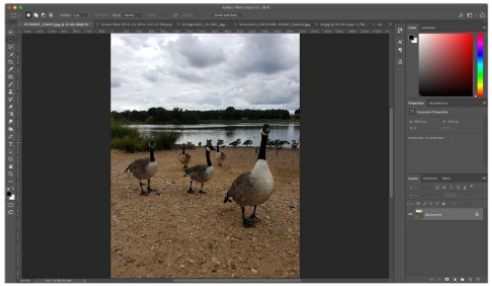


Image in Photoshop

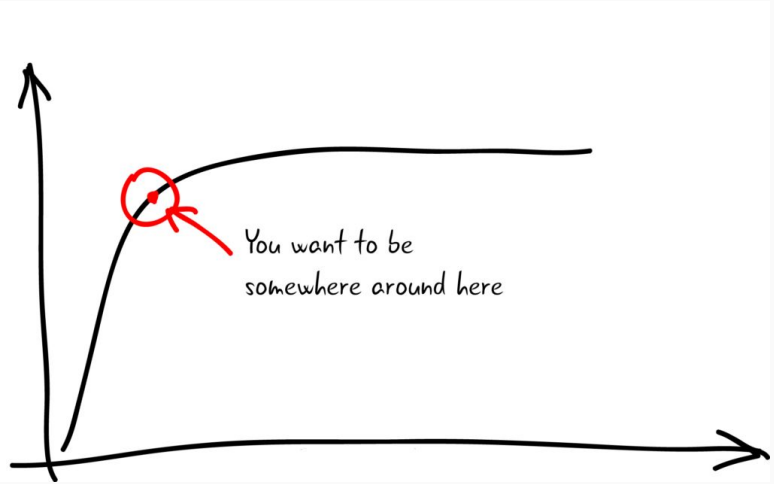
Camera

JPEG Pixel Data

Image in Photoshop

Optimize with a Goal in Mind

- Diminishing returns - if first 10% improvement takes x days, second 10% improvement might take 10x days.



- The goal should be a mix of product requirements and engineering feasibility (human level performance is often a good indicator).

Soft Labels

- Consider 'softening' your labels in cases where they are ambiguous or where you expect noise.

Labeller 1

Lion Tiger

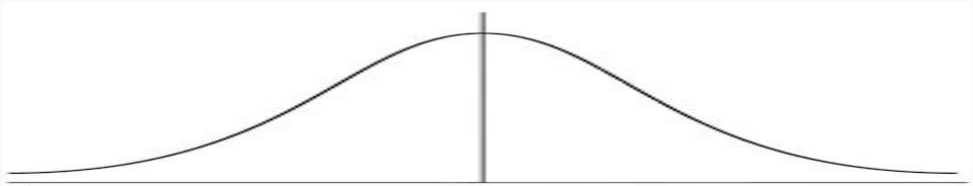
0	0	0	0	0	1	0	0	0	0

Labeller 2

0	0	0	0	0	0	1	0	0	0

Consider changing to

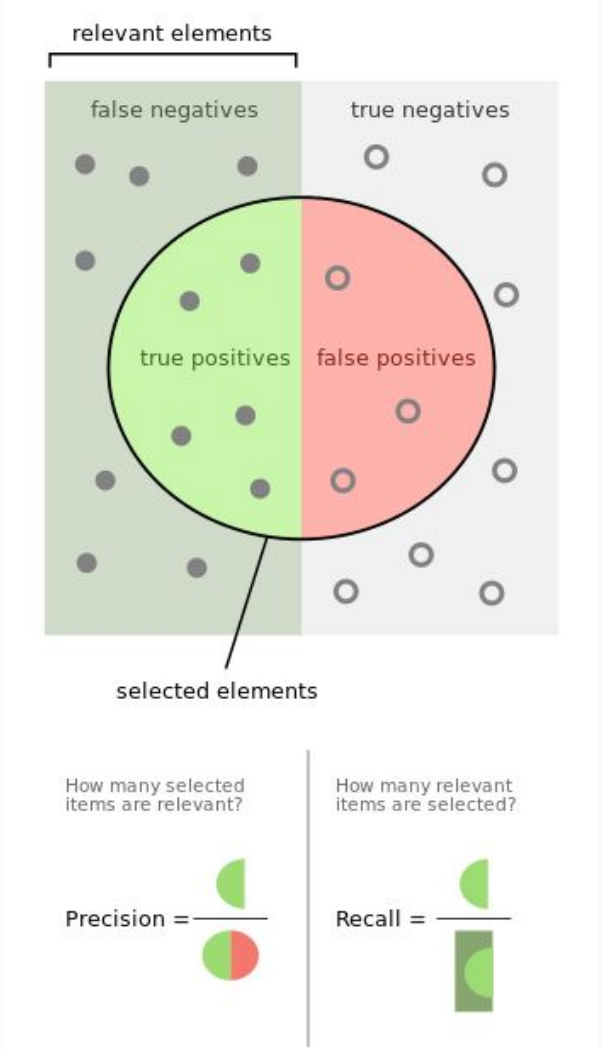
0.1	0.2	0.3	0.4	0.4	0.6	0.6	0.4	0.4	0.3



Get a Single Number Metric

- For example while looking at Precision or Recall separately is useful for debugging and getting intuition. For comparing models use **one number**.

- F1 score (harmonic mean of precision/recall)



Calibrate your model (For some applications)

Calibration: Prediction probabilities of the network are representative of the true correctness likelihood.

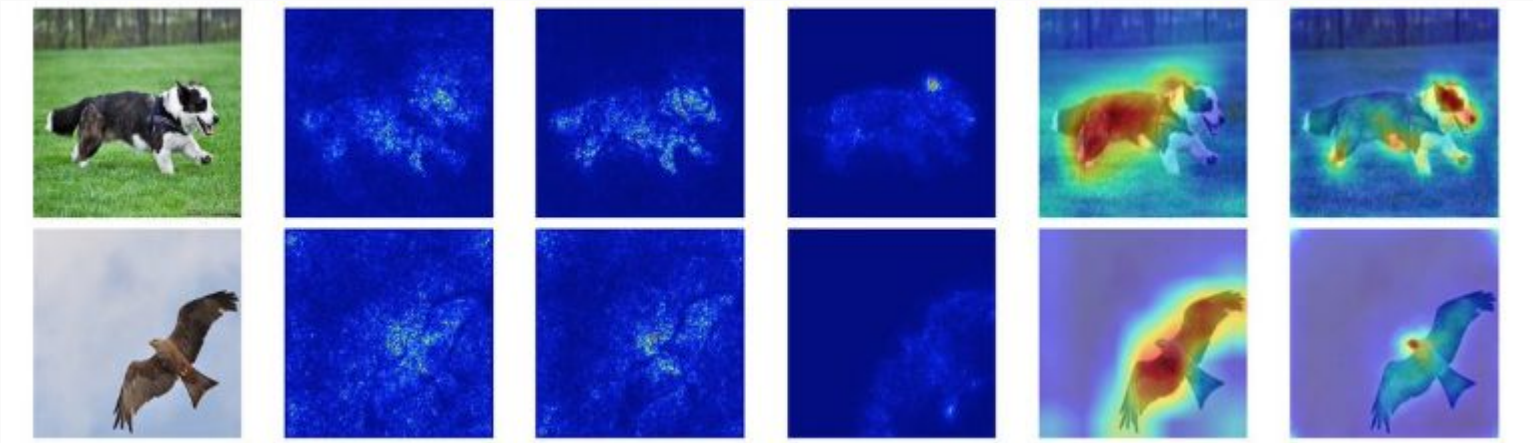
Paper: On Calibration of Modern Neural Networks

Temperature Scaling (one line change):

logits = logits / **T**

Consider Interpretability

- Hand crafted features lend themselves more to interpretability.
- Integrated gradients, saliency maps etc.
- Medical AI finding new indicators.



Gathering Data

https://xkcd.com/1838/

The possibilities

- Use an Existing Dataset
- Collecting your own data
  - Lots of sub-categories here.
- Unsupervised learning (unlabelled dataset).
- Adapting an existing dataset.

The possibilities

- **Use an Existing Dataset**
- Collecting your own data
  - Lots of sub-categories here.
- Unsupervised learning (unlabelled dataset).
- Adapting an existing dataset.

Using an Existing Dataset

https://toolbox.google.com/datasetsearch

The possibilities

- Use an Existing Dataset
- **Collecting your own data**
  - Lots of sub-categories here.
- Unsupervised learning (unlabelled dataset).
- Adapting an existing dataset.

Collecting your own data

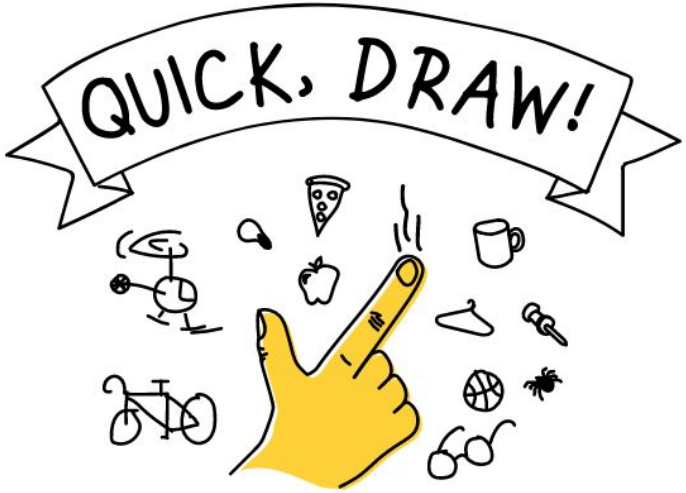
- From Scratch aka Crowdsourcing.

Crowdsourcing good practices:

- Use best of N ratings. (Cost / noise trade off)
- 'Best of N ratings' -> Soft labels
- Start off with a small sample of images and get feedback on edge cases.

Collecting your own data

Gamify!



Can a neural network learn to recognize doodling?

Help teach it by adding your drawings to the world's largest doodling data set, shared publicly to help with machine learning research.

Let's Draw!

Synthetic Data

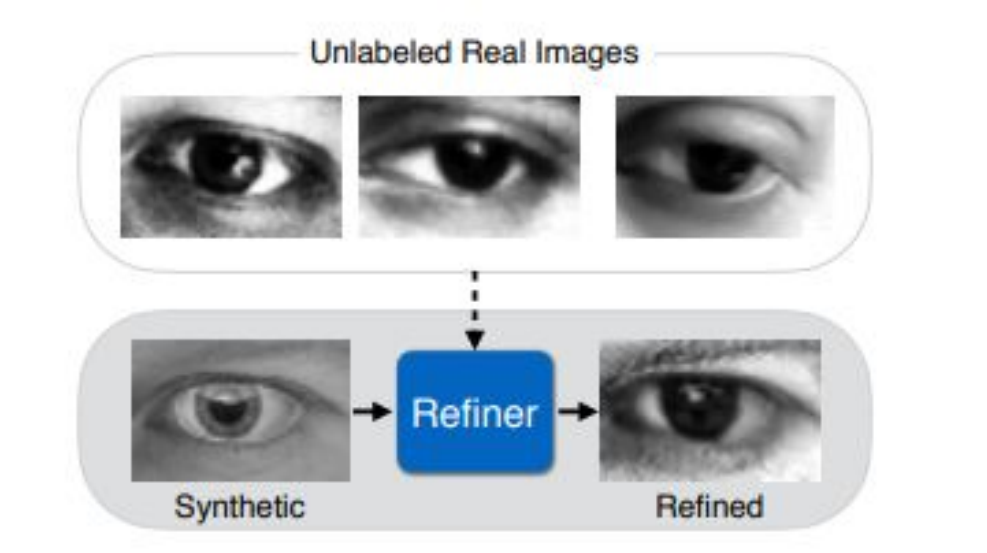
- Be careful of cross-domain problems and transferability.

Synthetic Data

- Fake it till you make it (Depth from stereo using synthetic data).

Synthetic Data

- [Learning from Simulated and Unsupervised Images through Adversarial Training]



Active Learning

- Guide: www.datacamp.com/community/tutorials/active-learning
- TL;DR Run an initial ('weakly trained') model on unlabelled data.
- Use results to select more data for training.
- Reduces time and data required for training.

Active Learning Examples

- Revisiting knowledge transfer for training object class detectors
- Large-scale interactive object segmentation with human annotators
- We don’t need no bounding-boxes: Training object class detectors using only human verification.

Utilizing Causal Relationships

- Use causal relations to narrow down the data your crowd computing workers have to sift through.
- Building a dataset for detecting questions? Questions usually start with *what, how , why, where*. Use this information to narrow down the text shown to human labelers.

Collecting your own data

- Get Creative with data augmentation
- Use language models to autocomplete sentences (GPT2, BERT are amazing).

I did not enjoy this movie. The film is big and beautiful but it is all over the place and it just does not flow. The characters are not funny or interesting enough to the point that I liked their jokes.

The possibilities

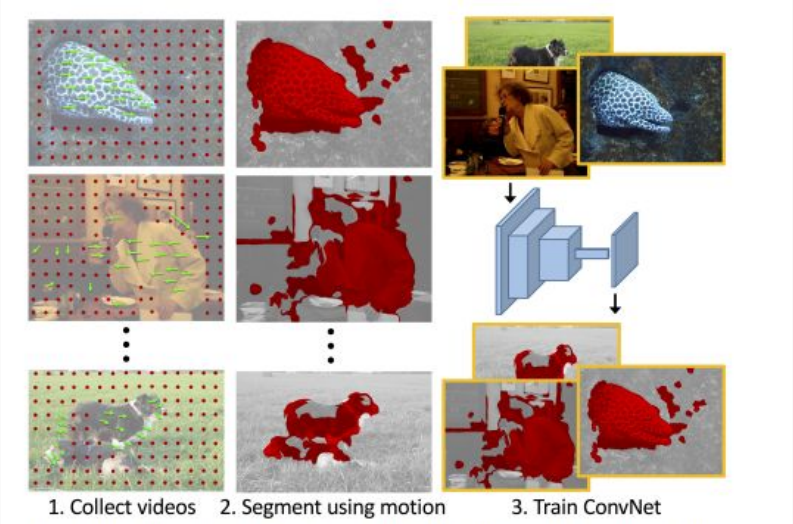
- Use an Existing Dataset
- Collecting your own data
  - Lots of sub-categories here.
- **Unsupervised learning.**
- Adapting an existing dataset.

Gameplay

- Self playing agents (video games, board games)

Unsupervised Learning

- Not only for RL problems.
- Learning Features By Watching Objects Move (Optical flow + convnet)

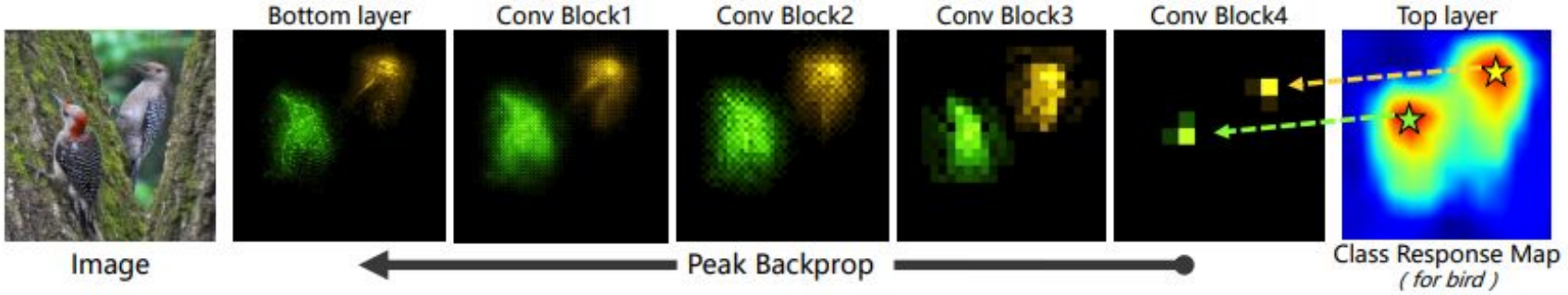


The possibilities

- Use an Existing Dataset
- Collecting your own data
  - Lots of sub-categories here.
- Unsupervised learning (unlabelled dataset).
- **Adapting an existing dataset.**

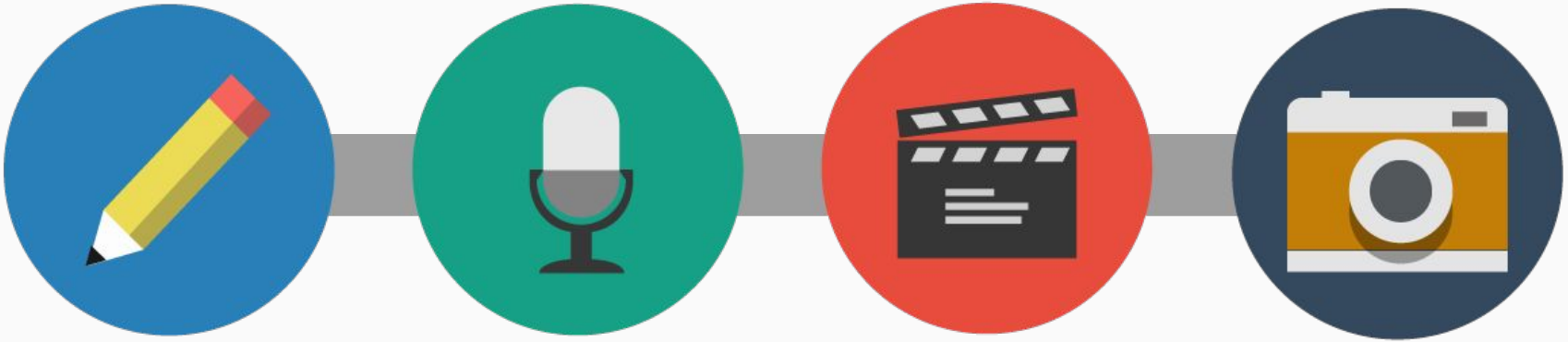
Adapting an existing dataset

- Existing dataset close but not exactly what you need? Adapt!
- Example: Segmentation from bounding boxes.
- **Weakly Supervised Instance Segmentation using Class Peak Response**



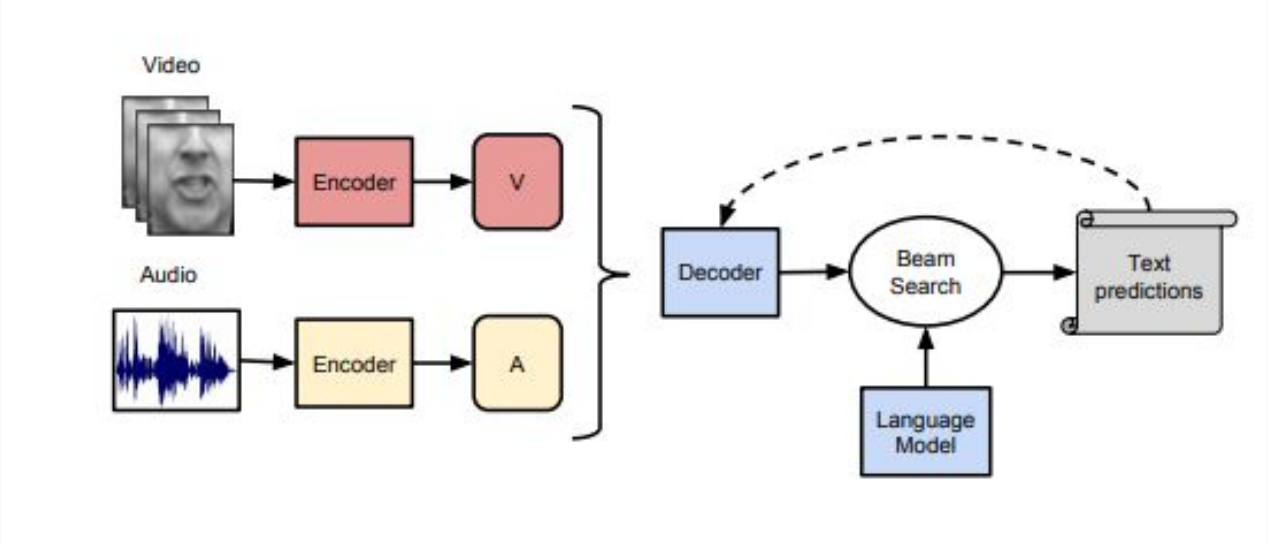
Go Multimodal

Some tasks are easier when you have two or more modalities lined up (audio, video and text).



Go Multimodal

Deep Audio-Visual Speech Recognition



The cost/time/performance tradeoff

At the end it's a tradeoff between time, quality, ease of use.

Compare the cost of SWE-hours vs data collection and model training costs to decide.

Talking to your dataset

Balancing

- Undersample most common classes.
- Oversample (with augmentation) under represented classes.
- Weighted loss (with weight by class).

Visualize Your Data

Great to get a feel for the data.

Randomly select some samples and visualize them, just to spot check for accuracy.

Visualize Your Data

Fun ways of visualization. (Does not have to be just listening to audio or looking at pictures). Sentiment Classification Visualization.

Apache Beam

1. https://beam.apache.org/
2. Data processing (streaming as well as batch) pipeline.
3. Simple python usage -> Write function for one data element, map and distribute over entire dataset.


That's all Folks!