WHO IS ALYCE?

Alyce.ai™ is the Predictive AI Platform that converts subtle data patterns into actionable and scalable insights.

- Automate data cleansing, classification, and model selection
- Leverage sophisticated machine learning and AI to understand what is impacting your business
- Unlock your competitive advantage by transforming subtle data patterns into valuable insights

OBJECT COMPUTING
HOME TO ALYCE

Mythbusters Episode #1:

I Need Tons of Data to Implement Machine Learning

Meet the Speakers

Tonya Ehlmann

Jason Bull

Xiao Yang

Do I really need lots of data to build an optimized model?

Buzz on the street used to be …

“I need tons of data”

&

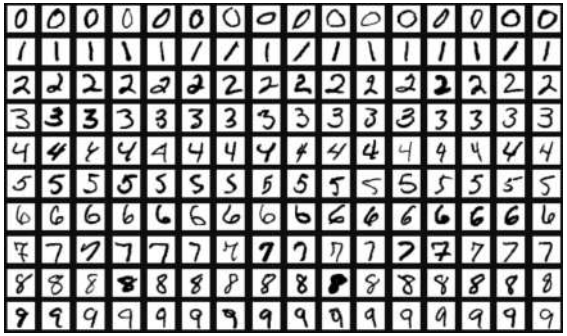
“The more data I have, the better off I will be”

Repeating pattern proven effective ...

Clear business outcomes first → Right size data → Validate & predict

What you learn in a classroom ...

Original MNIST: 70K

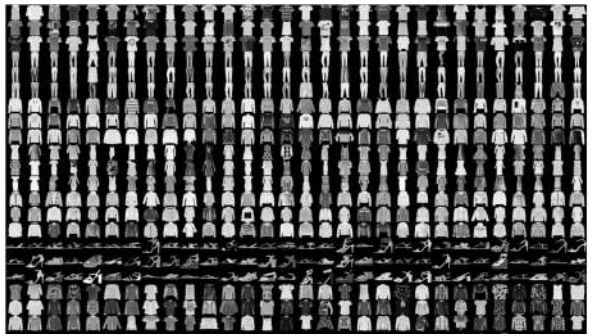


For "average" machine learning problems, your training set size is roughly 10x the number of free parameters in your model.

Yaser Abu-Mostafa, Caltech

We need at least 1,000 images per class to train a good object-detection engine.

FASHION-MNIST: 70K



Our experience with different industry verticals ...

"I don't have lots of data; how should I think about implementing ML?"

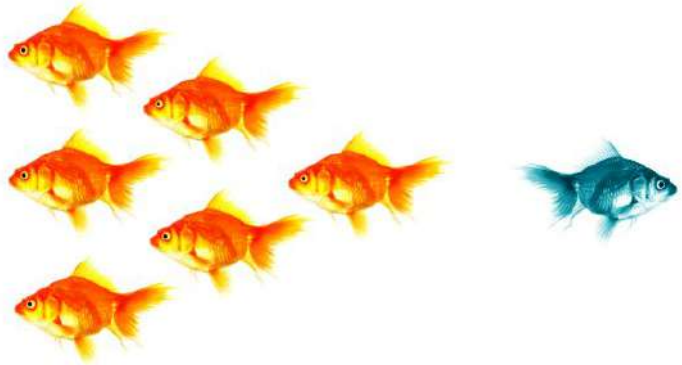
"I have too much data; I don't know how to make sense of it …"

"More than three quarters of large companies today have a "data-hungry" AI initiative underway. Yet, many of the most valuable data sets in organizations are quite small: Think kilobytes or megabytes rather than exabytes."

Harvard Business Review, 2020

"(How to train models with less data) it's a challenge, it's a goal, and there's certainly reason to believe that it's possible."

Rob High, IBM Watson CTO, 2018



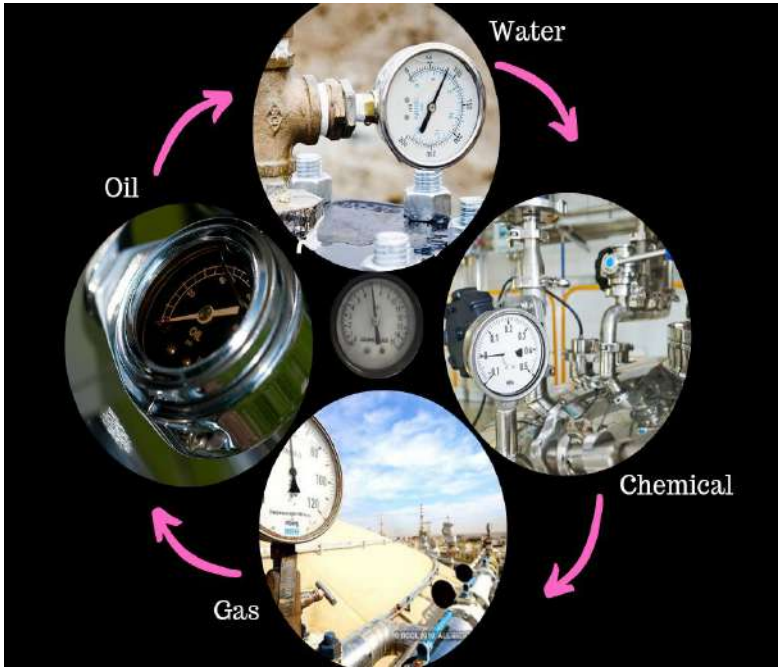
New thinking.

New approach.

New outcomes.

Use Case #1: Building a deep neural network with ONE data point

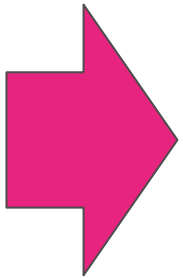
Analog gauges still very popular



Inconsistent readings happen all the time!

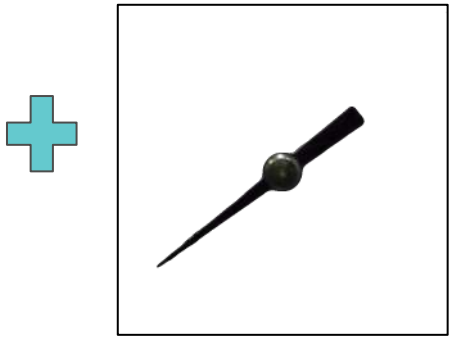
Our deep learning objective started with one data point.

Any analog gauge

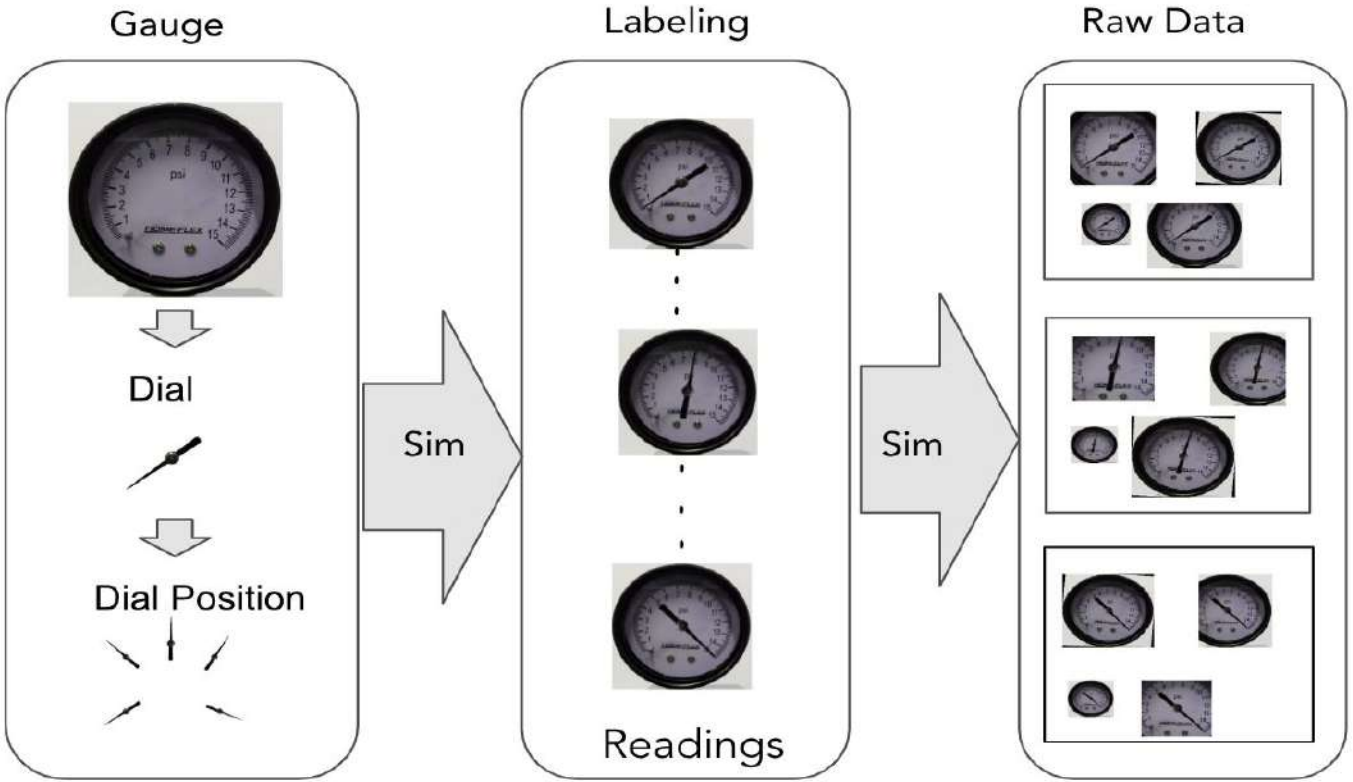


Face

Dial



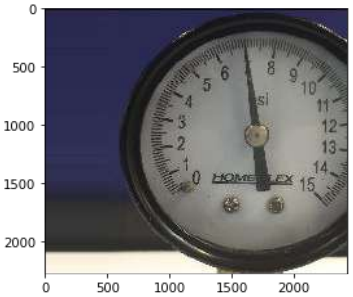
We simulated some images to enable initial model training.



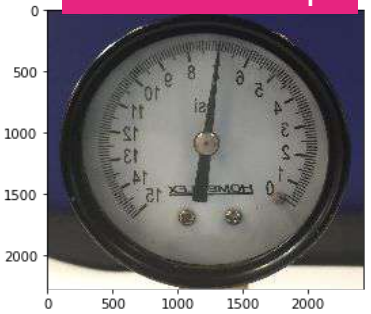
We further augmented data during the training stage.

Original

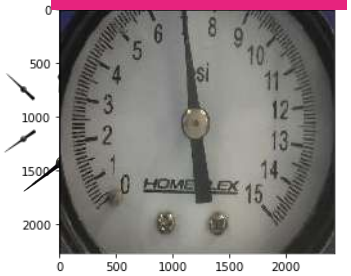
Width Shift



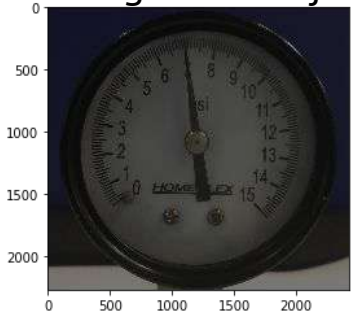
Horizontal Flip



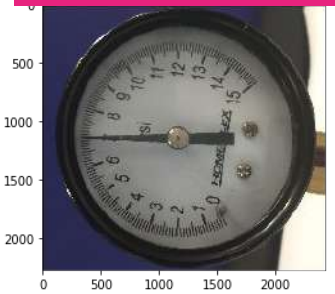
Random Zoom



Brightness Adj

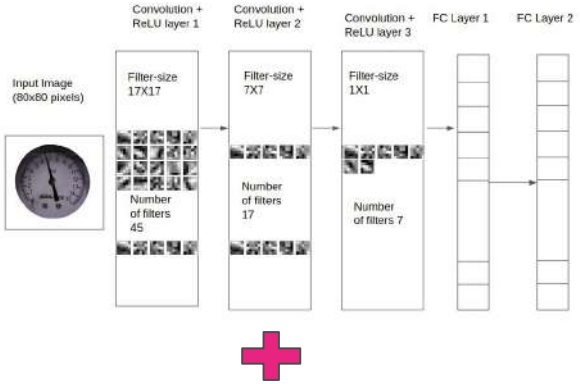


Random Rotation

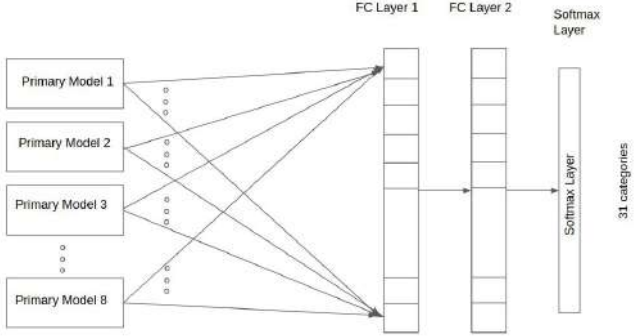


We can build a robust and generic model with high accuracy.

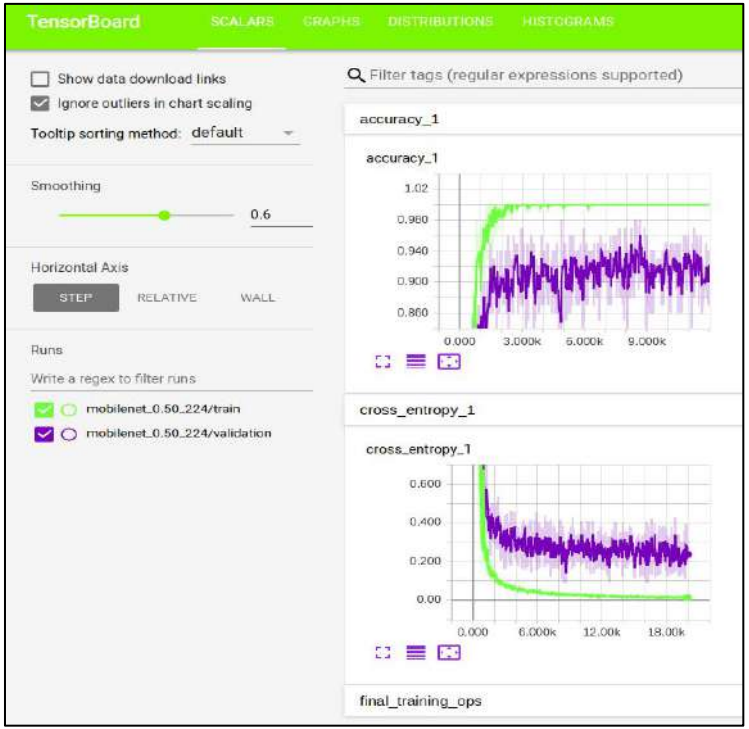
CNN Architecture: Primary



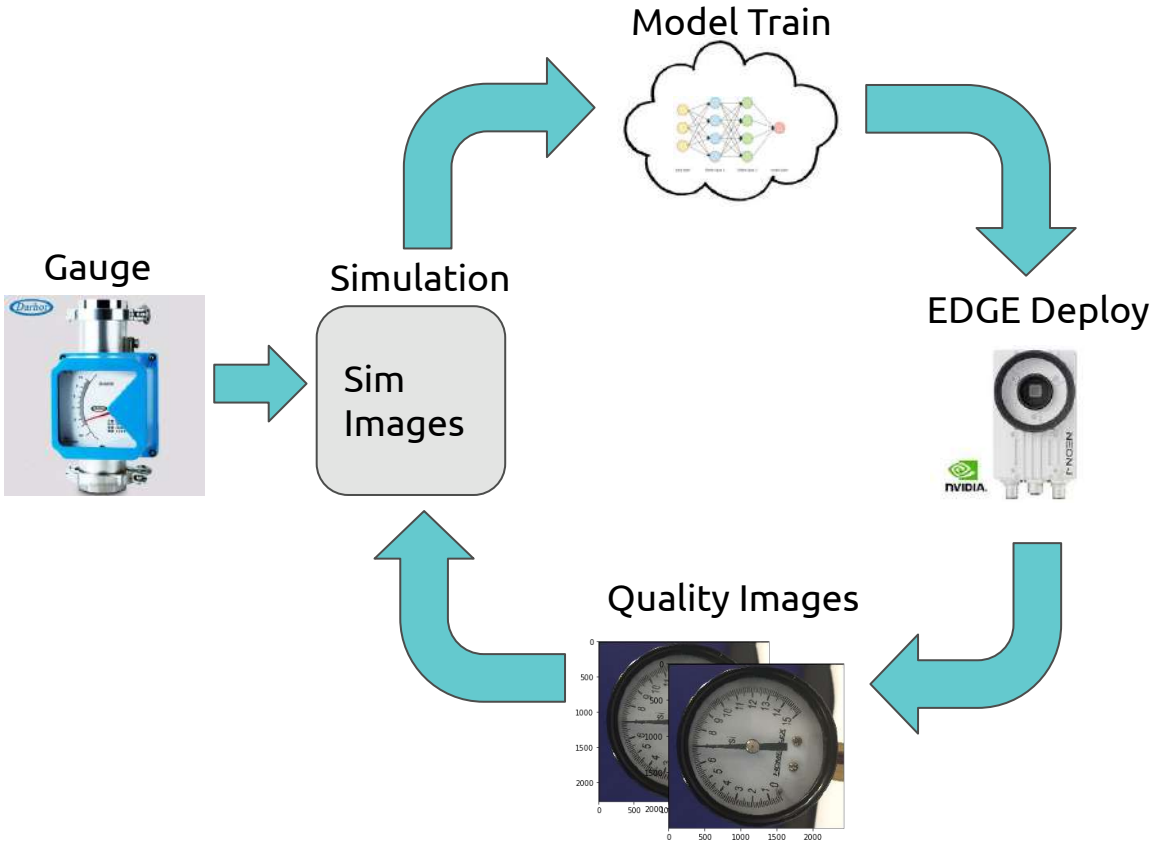
Super Learner: Ensemble



High Accuracy achieved (>90% acc)



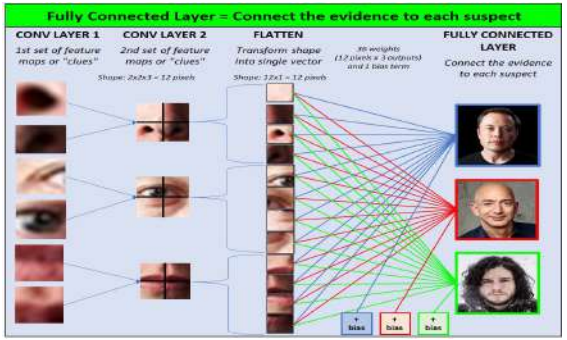
Validation framework needed to improve the models iteratively

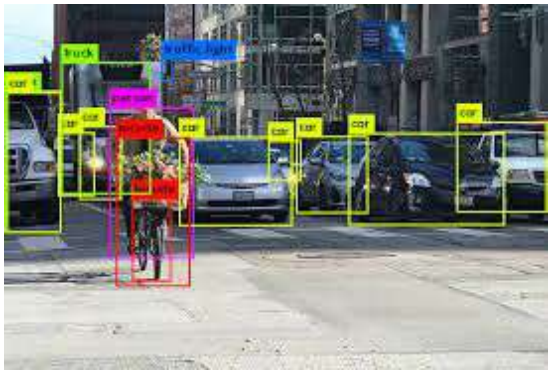


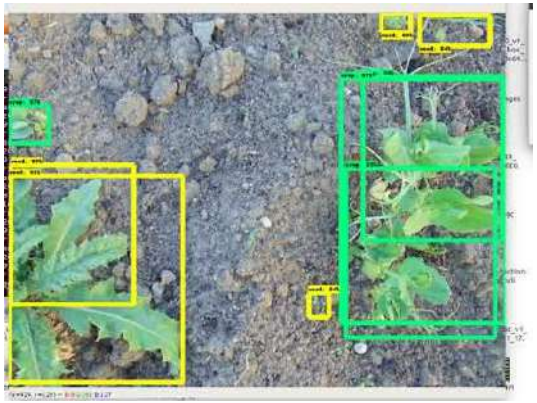
Takeaways

- Simulation helps
- Usable model fast … initial value
- Improving and expanding data is valuable
- Iteratively validate and refine your predictions

Use case #2: Build an image-based object detection model with some data







I need a model to detect the subtle differences among 3 dog breeds, but my model struggles ...

Brittany Spaniel

50 images

Clumber Spaniel

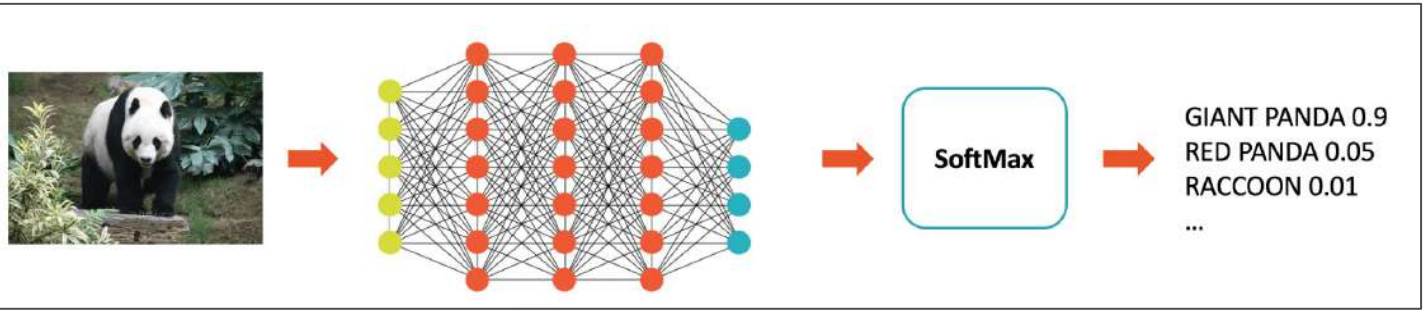
40 images

Golden Retriever

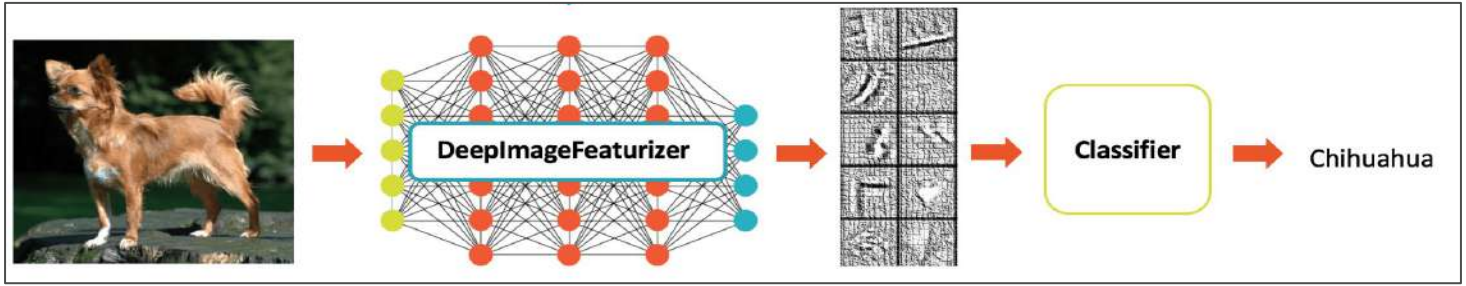
1200 images

Transfer learning applied to model training, fewer images needed

Pre-trained DNN models (Inception V3, Xception, ResNet, etc.)

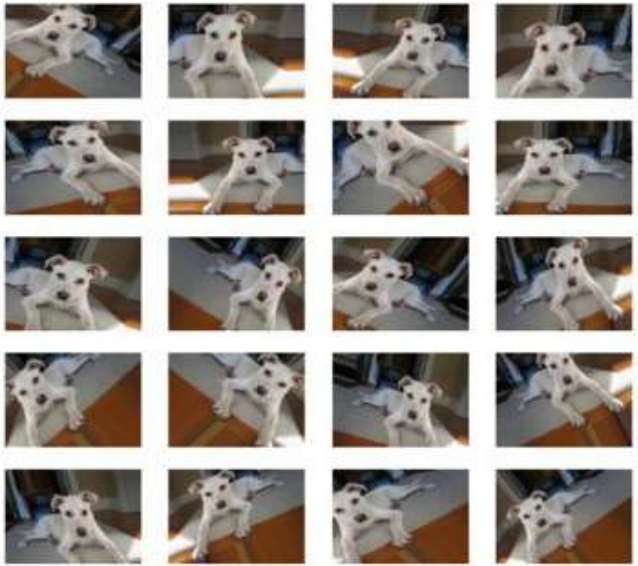


Custom-trained DNN models for Breed Detection

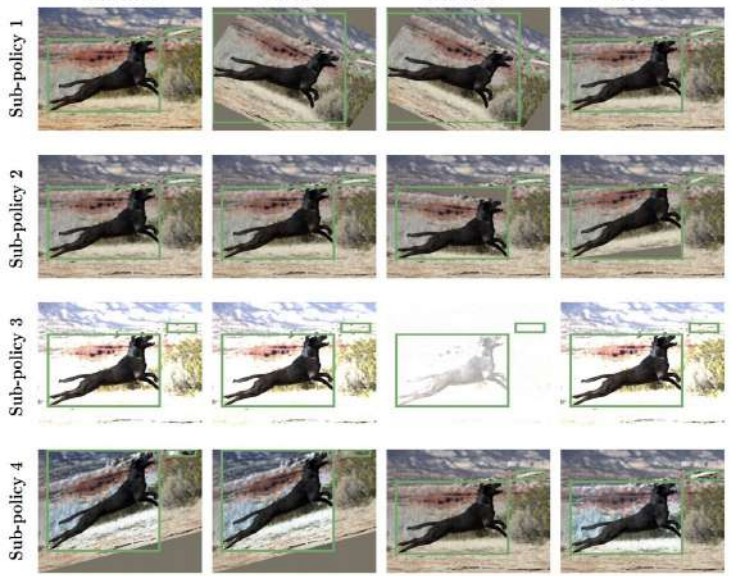


Getting more training images through optimized data augmentation strategies

Data Augmentations

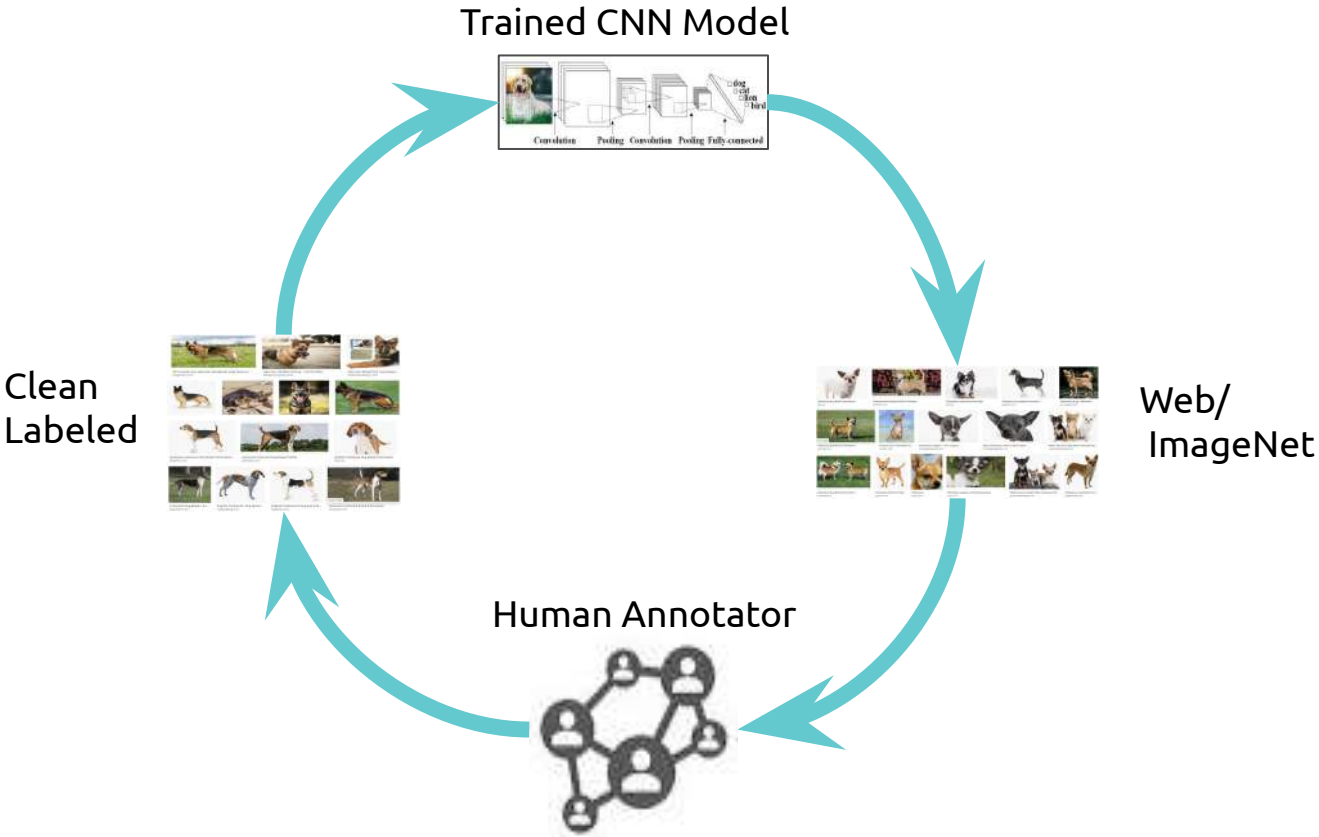


Optimized augmentation strategies



Google Research, 2019

Smart crowd sourcing: More labeled images can be sourced through active learning.

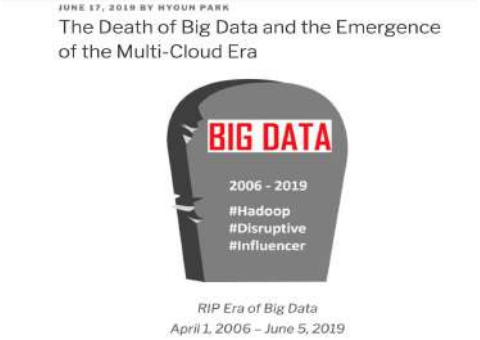


Takeaways

- Leveraging pre-trained architectures can improve predictions
- Data augmentation strategy can be optimized to improve
- Semi-supervised classification can accelerate performance
- Iteratively validate and refine your predictions

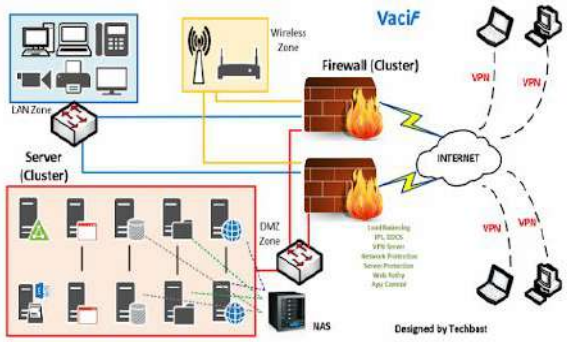
Use Case #3: Building a valuable ML model with lots of data (Big Data)

Is big data really dead?



Formulating your business outcome first is the key to the success of your big data project.

Daily TB to Monitor Anomalies



Events recorded in real time...
Reactive to “bad events”
$ spend on support resource

Formulation of your ML Project

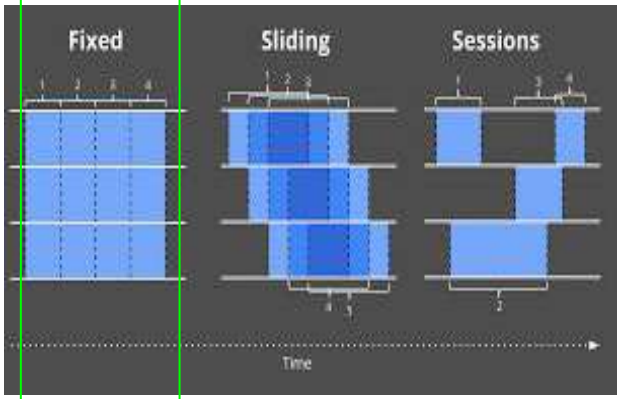
- What are the most important events?
- How much resource do you spend on those?
- Can we predict “critical events” ahead of time?
- If your model is imperfect, is it still useful?

Reducing your “Big Data” to the right size is critical.

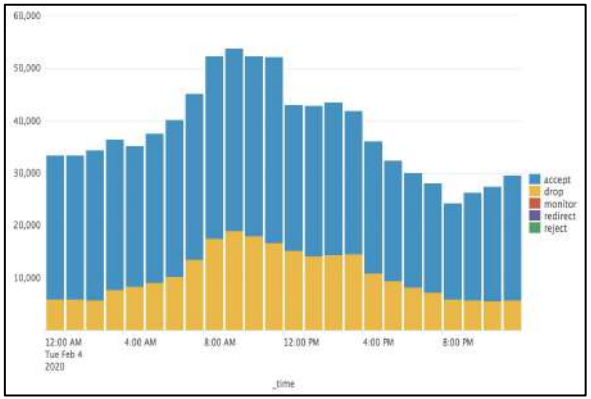
Raw events in real time

```
time=1573755736
|loc=755834
|fileid=1573711200
|action=drop
|orig=321.12.34.56
|i/f_dir=inbound
|i/f_name=eth2
|has_accounting=0
|logId=-1
|log_type=log
|log_sequence_num=0
|is_first_for_luuid=131072
|log_version=1
|uuid=<00000000,00000000,00000000,00000000>
|product=VPN-1 & FireWall-1
|rule=23
|rule_uid={70732248-9C25-470D-9562-6EDA01F91B77}
|rule_name=Drop rule
|src=375.243.612.160
|s_port=46959
|dst=22.55.77.99
|service=8080
|proto=tcp
|__policy_id_tag=product=VPN-1 & FireWall-1[db_tag={4E6
|origin_sic_name=CN=STL_DEVICE_23,O=CMA-PRI..99qqct
```

Aggregate through “windowing”



Id interesting patterns



Takeaways

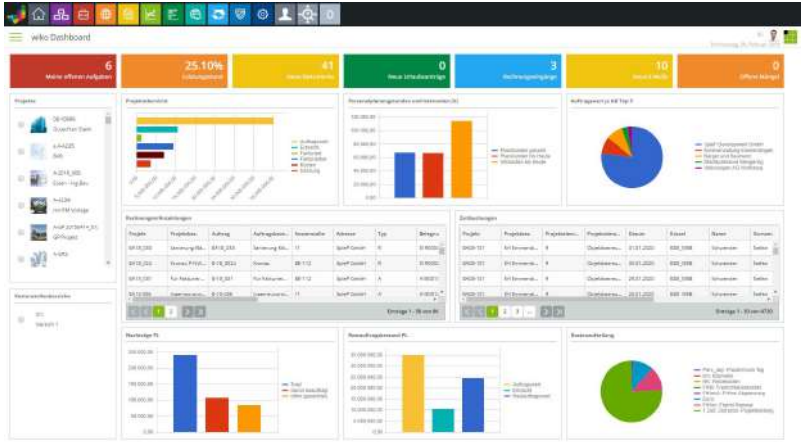
- Naming the business outcome is important
- Featuring can massively reduce data size and velocity
- Iteratively validate and refine your predictions

These core principles are valuable to the success of your ML projects.

Clear business outcomes first → Right size data → Validate & predict

Next Myth to Bust:

What isn't your dashboard telling you?



ALYCE™

ALYCE.AI

Follow me on LinkedIn and Twitter to stay up to date on the intersection of AI and business.

https://www.linkedin.com/showcase/alyce.ai/

https://twitter.com/alyce_ai