Computer Vision & ML Model Tuning Workshop

The University of Texas at Dallas

4/5/2025 - 4/6/2025

“Safe Harbor” Statement

In this presentation we have made forward-looking statements. These statements are based on our estimates and assumptions and are subject to risks and uncertainties. Forward-looking statements include the information concerning our possible or assumed future results of operations. Forward-looking statements also include those preceded or followed by the words "anticipates," "assumes," "believes," "estimates," "expects," "forecasts," "hopes," "intends," "plans," "targets" or similar expressions. For those statements, we claim the protection of the safe harbor for forward-looking statements contained in the Private Securities Litigation Reform Act of 1995. We undertake no obligation to revise or publicly release the results of any revision to these forward-looking statements, except as required by law. Given these risks and uncertainties, readers are cautioned not to place undue reliance on such forward-looking statements. The following important factors, along with those discussed in our filings with the Securities and Exchange Commission (the "SEC"), could affect future results and could cause those results to differ materially from those expressed in the forward-looking statements: the effects of competition in the markets in which we operate, including the inability to successfully respond to competitive factors such as prices, promotional incentives and evolving consumer preferences; failure to take advantage of, or respond to competitors' use of, developments in technology, including artificial intelligence, and address changes in consumer demand; performance issues or delays in the deployment of our 5G network resulting in significant costs or a reduction in the anticipated benefits of the enhancement to our networks; the inability to implement our business strategy; adverse conditions in the U.S. and international economies, including inflation and changing interest rates in the markets in which we operate; cyberattacks impacting our networks or systems and any resulting financial or reputational impact; damage to our infrastructure or disruption of our operations from natural disasters, extreme weather conditions, acts of war, terrorist attacks or other hostile acts and any resulting financial or reputational impact; disruption of our key suppliers' or vendors' provisioning of products or services, including as a result of geopolitical factors or the potential impacts of global climate change; material adverse changes in labor matters and any resulting financial or operational impact; damage to our reputation or brands; the impact of public health crises on our business, operations, employees and customers; changes in the regulatory environment in which we operate, including any increase in restrictions on our ability to operate our networks or businesses; allegations regarding the release of hazardous materials or pollutants into the environment from our, or our predecessors', network assets and any related government investigations, regulatory developments, litigation, penalties and other liability, remediation and compliance costs, operational impacts or reputational damage; our high level of indebtedness; significant litigation and any resulting material expenses incurred in defending against lawsuits or paying awards or settlements; an adverse change in the ratings afforded our debt securities by nationally accredited ratings organizations or adverse conditions in the credit markets affecting the cost, including interest rates, and/or availability of further financing; significant increases in benefit plan costs or lower investment returns on plan assets; changes in tax laws or regulations, or in their interpretation, or challenges to our tax positions, resulting in additional tax expense or liabilities; changes in accounting assumptions that regulatory agencies, including the SEC, may require or that result from changes in the accounting rules or their application, which could result in an impact on earnings; and risks associated with mergers, acquisitions, divestitures and other strategic transactions, including our ability to consummate the proposed acquisition of Frontier Communications Parent, Inc. and obtain cost savings, synergies and other anticipated benefits within the expected time period or at all.

As required by SEC rules, we have provided a reconciliation of the non-GAAP financial measures included in this presentation to the most directly comparable GAAP measures in materials on our website at www.verizon.com/about/investors.

What is Computer Vision?

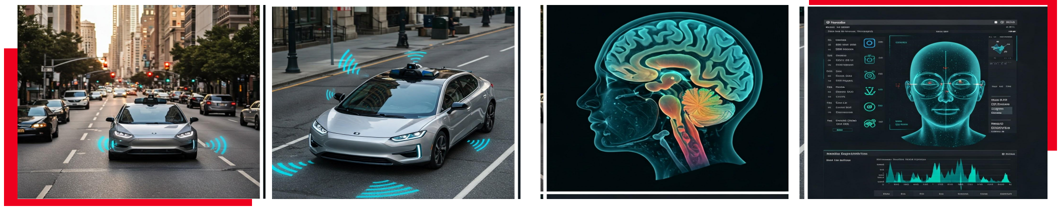
Field of AI that enables computers to interpret and understand visual world. Computer vision technique to locate and identify objects in images/videos.

Gain high-level understanding from images or videos. Extracting meaning and insights from visual data (images, videos). Determines object's position, size and class (e.g. car, person, traffic light)

Key Applications:

- **Object detection and recognition**
- **Image classification and segmentation**
- **Facial recognition and analysis**
- **Medical imaging and diagnosis**
- **Autonomous vehicles**
- **Number plate reading**

Source: Gemini image generation



Source: Source: Gemini image generation

Core Concepts in Computer Vision

Image Processing

Techniques for manipulating & analyzing digital images.

- Image transformation modifies or changes images
- Image enhancement improves clarity and reduces distortions
- Noise reduction preserves image features and texture
- Morphological operations analyze images by structure

Techniques enhance and analyze digital images by modifying, clarifying, and structuring them

Feature Extraction

Relevant features are identified & extracted from raw data.

- Edge detection identifies object boundaries via intensity or color changes
- Corner detection finds distinctive points stable across image changes
- Feature descriptors create compact representations of keypoints for image matching

Identifies and isolates key elements, such as edges and corners, to facilitate machine learning analysis

Deep Learning

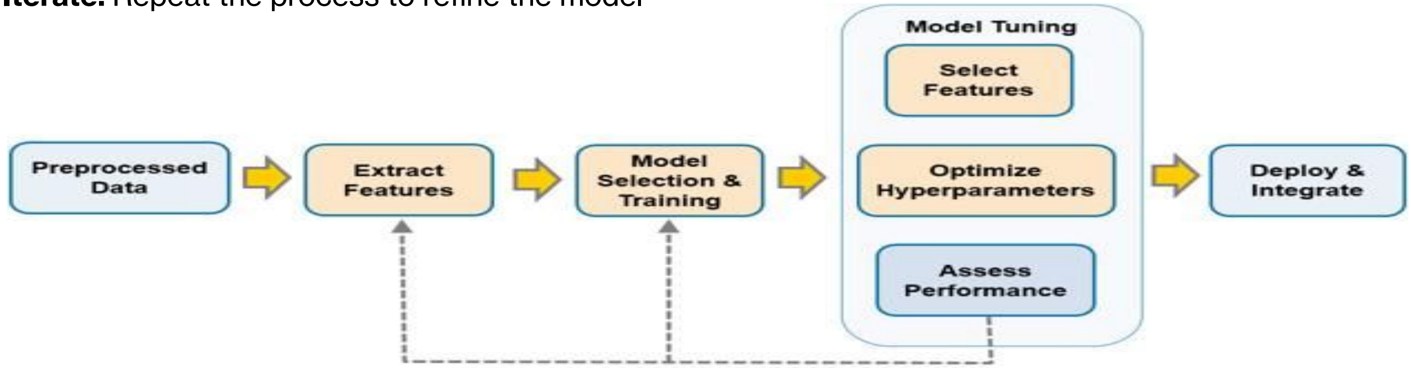
Allows machines to interpret visual data in unprecedented ways.

- CNN(Convolutional Neural Networks)s learn spatial hierarchies of image features
- GANs use two networks to create realistic images
- VAEs learn distribution over latent space
- ViTs process images using self attention mechanisms
- Vision Language Models integrate visual and textual data

Deep learning enables machines to interpret visual data through advanced models like CNNs, GANs, and ViTs.

The ML Model Tuning Process

- **Data Collection and Preparation:** Gather and preprocess relevant data
- **Model Selection:** Choose an appropriate model architecture
- **Hyperparameter Tuning:** Optimize model parameters
- **Training:** Train the model on the prepared data
- **Evaluation:** Assess model performance using metrics
- **Iterate:** Repeat the process to refine the model



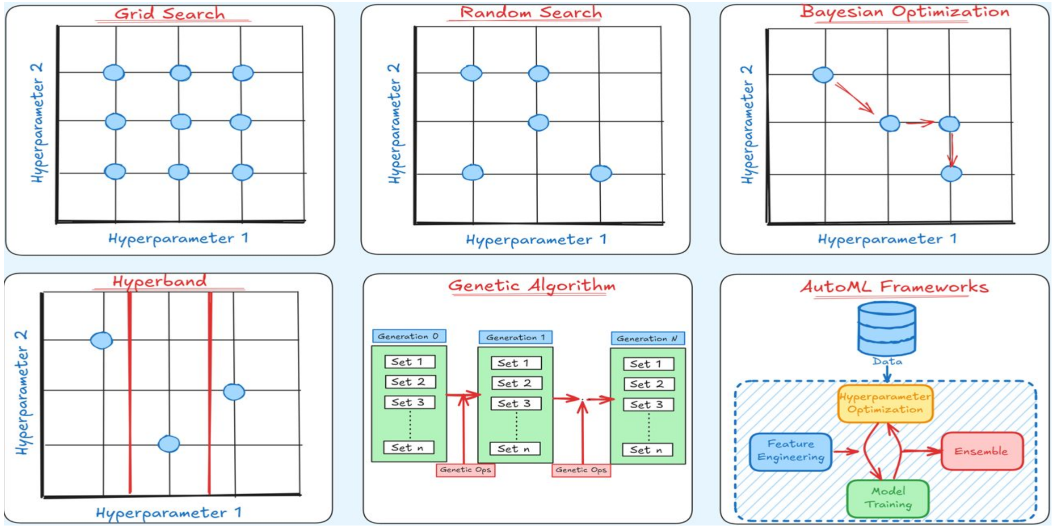
https://blogs.mathworks.com/loren/2020/06/13/building-optimized-models-in-a-few-steps-with-automl/

Hyperparameter Optimization Techniques

Siva

Hyperparameter tuning is the process of optimizing the hyperparameters of a machine learning model to enhance its performance.

- **Grid Search:** Exhaustively search all parameter combinations
- **Random Search:** Sample a subset of parameter space randomly
- **Bayesian Optimization:** Build a probabilistic model of the objective function



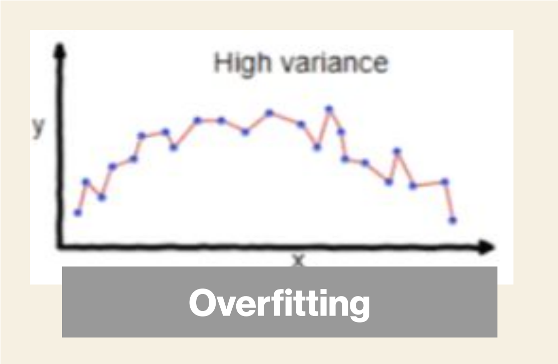
Source:
https://www.nb-data.com/p/6-common-hyperparameter-optimization

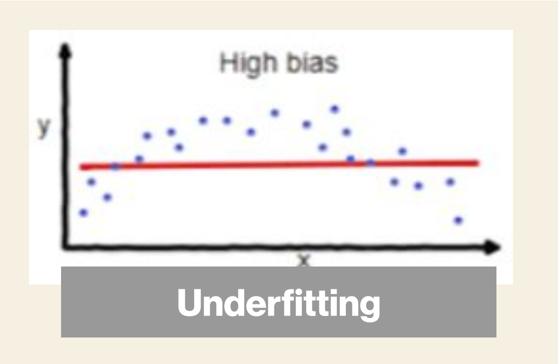
Regularization Methods

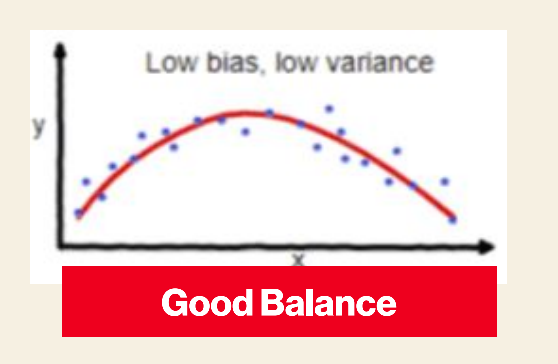
Regularization prevents overfitting by penalizing model complexity, thus encouraging simpler, more general models that avoid extreme predictions caused by noise.

Following are the Regularization Techniques:

1. Lasso Regularization – (L1 Regularization)
2. Ridge Regularization – (L2 Regularization)
3. Elastic Net Regularization – (L1 and L2 Regularization)







Source:https://www.geeksforgeeks.org/regularization-in-machine-learning/

Data Augmentation Strategies

Data augmentation is a technique of artificially increasing the training set by creating modified copies of a dataset using existing data. It includes making minor changes to the dataset or using deep learning to generate new data points.

When should you use data augmentation?

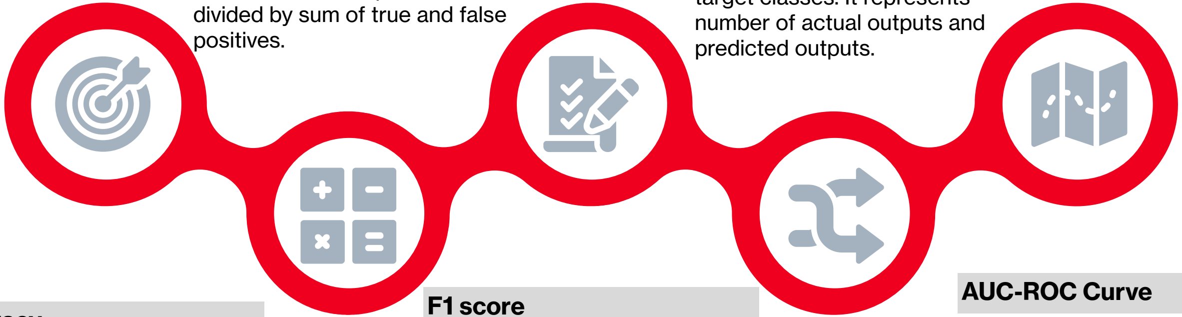
- To prevent models from overfitting
- The initial training set is too small
- To improve the model accuracy
- To Reduce the operational cost of labeling and cleaning the raw dataset

Audio Data Augmentation	Text Data Augmentation	Image Augmentation
• **Noise injection:** add random noise to enhance model performance. • **Shifting:** time-shift audio to simulate varied playback. • **Changing speed:** adjust audio tempo for data variation. • **Changing pitch:** modify audio pitch for diverse samples.	• **Word/sentence shuffling:** Randomly reorder words or sentences. • **Word replacement:** Substitute words with synonyms. • **Syntax-tree manipulation:** Paraphrase sentences while retaining core words. • **Random word insertion:** Add random words to the text. • **Random word deletion:** Remove random words from the text.	• **Geometric transformations:** Randomly flip, crop, rotate, stretch, or zoom images; avoid excessive combinations to maintain performance. • **Color space transformations:** Adjust RGB channels, contrast, and brightness randomly. • **Kernel filters:** Randomly alter image sharpness or blurring. • **Random erasing:** Delete portions of the original image. • **Mixing images:** Blend and combine multiple images.
Essentially, we're making the audio sound a bit different to teach the model to recognize it in various real-world scenarios.	**"Basically, we're teaching the model to understand the same meaning, even when the words are all mixed up!"**	**"We're essentially showing the model the same thing in a million different ways, so it truly gets it."**

Siva

Evaluating Model Performance

Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Model is evaluated based on the following metrics:



Accuracy

Accuracy measures the ratio of correct predictions to total predictions, serving as a fundamental model evaluation metric.

Precision and Recall

Precision measures the accuracy of positive predictions, calculated as true positives divided by sum of true and false positives.

F1 score

The F1 score, the harmonic mean of precision and recall, balances the trade-off between them, combining both metrics into a single measure.

Confusion Matrix

Confusion matrix is a N x N matrix where N is the number of target classes. It represents number of actual outputs and predicted outputs.

AUC-ROC Curve

AUC (Area Under Curve) analyzes classification model performance across thresholds, while the ROC curve visually represents this performance probabilistically.

Confusion Matrix Terms

- **True Positive (TP):** Actual and predicted values are both YES.
- **True Negative** (TN): Actual and predicted values are both NO.
- **False Positive** (FP): Actual is NO, predicted is YES.
- **False Negative** (FN): Actual is YES, predicted is NO.

AUC-ROC Curve Patterns

- TPR (Recall): True Positive Rate.
- FPR: False Positive Rate (False Positives / False Positives + True Negatives).

Python libraries to check metrics

```
from sklearn.metrics import precision_score,\
recall_score, f1_score, accuracy_score
```

Source: https://www.geeksforgeeks.org/machine-learning-model-evaluation/

Popular Model : YOLO - Computer Vision and ML Model Tuning

- You Only Look Once (YOLO) is a state-of-the-art, real-time object detection system.
- YOLO divides the input image into a grid and simultaneously predicts bounding boxes and class probabilities for each grid cell.
- YOLO can be optimized by adjusting anchor boxes, learning rate, batch size and data augmentation.
- Model size and input resolution are other areas for optimization.

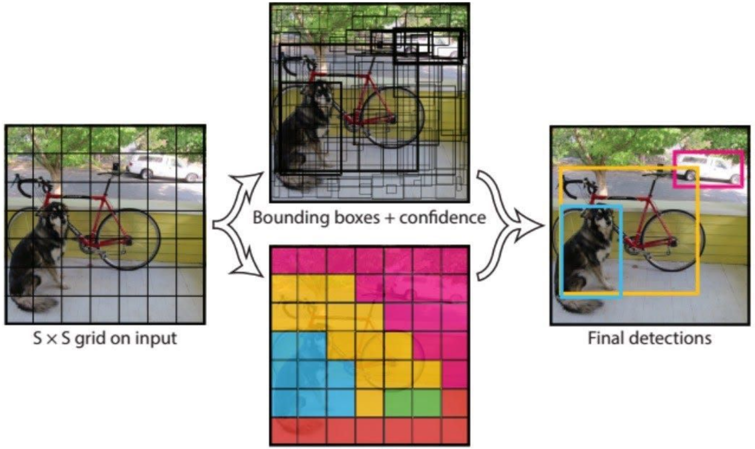


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

Source:https://iaee.substack.com/p/yolo-intuitively-and-exhaustively verizon