Machine Learning

2024 (ML-2024)

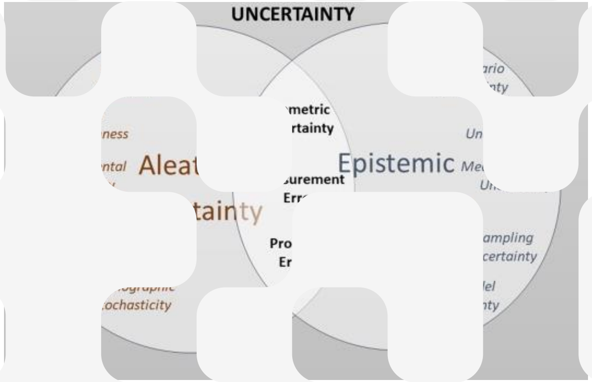
Lecture 8. Segmentation and object detection. Tips and tricks in ML.

by Alexei Valerievich Kornaev, Dr. habil. in Eng. Sc.,
Researcher at the RC for AI, Assoc. Prof. of the Robotics and CV Master’s Program, Innopolis University
Researcher at the RC for AI, National RC for Oncology n.a. NN Blohin
Professor at the Dept. of Mechatronics, Mechanics, and Robotics, Orel State University

ML-2024
by Alexei Kornaev

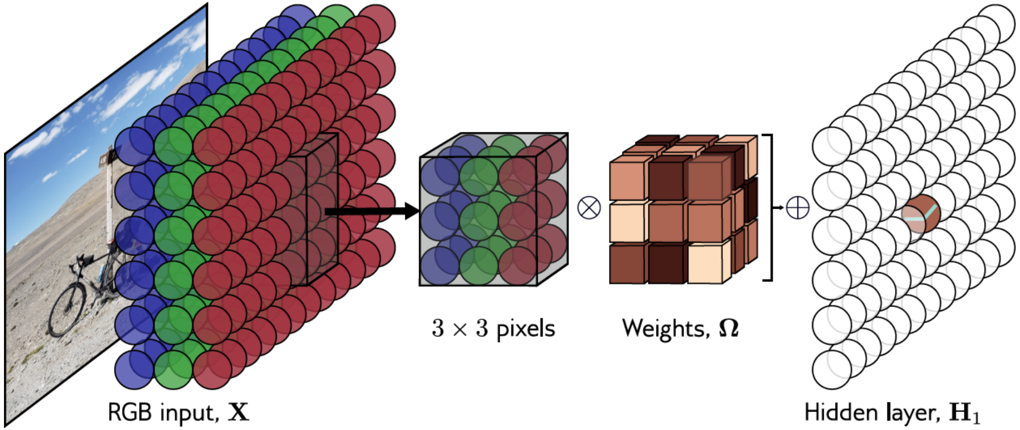
Agenda

I. RECAP ON CNNs
II. SEMANTIC SEGMENTATION
III. OBJECT DETECTION
IV. TIPS AND TRICKS IN ML



All models are wrong, but ~~some~~ models that know when they are wrong, are useful
/George Box + unknown researcher from the Google AI Brain Team/

CNN intuition

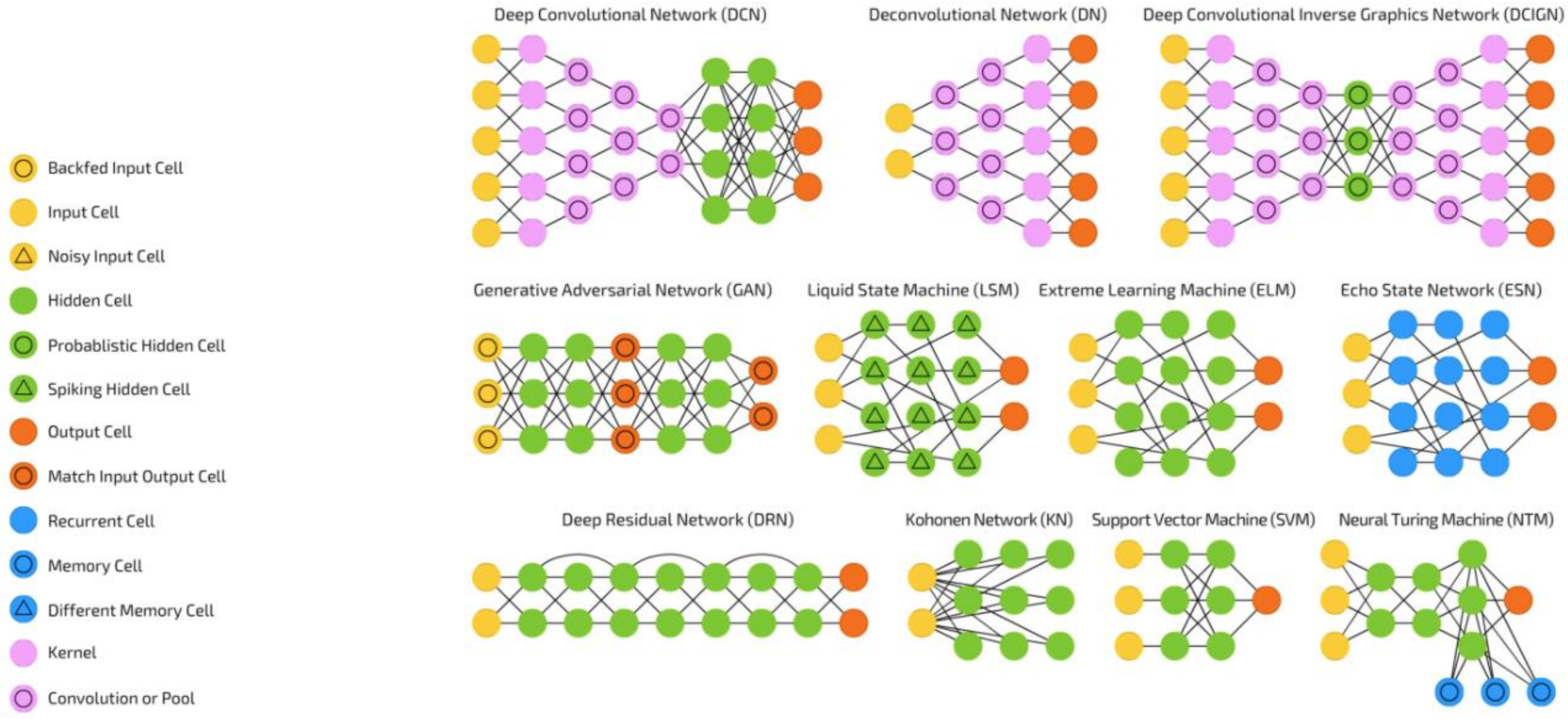


Algorithm:

1. Initialize weights $\boldsymbol{\Theta}^{(k)}$ randomly.
2. Calculate $\nabla L = \left[\partial L / \partial \theta_{ij}^{(k)}\right]$ with backpropagation.
3. Update weights $\boldsymbol{\Theta}^{(k)} : {\theta_{ij}^{(k)}}^{H} = {\theta_{ij}^{(k)}}^{C} - \alpha \frac{\partial L}{\partial \theta_{ij}^{(k)}}$.
4. Repeat pp. 2-3 until $L^{H} - L^{C} < \delta$ or #iter > N_{max}.
5. Save the best model (with min. validation loss): $\boldsymbol{\Theta}^{(k)}$.

S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

Recap: some of the ANN architectures



[Almost complete chart of NNs (2016)]

ML-2024
by Alexei Kornaev

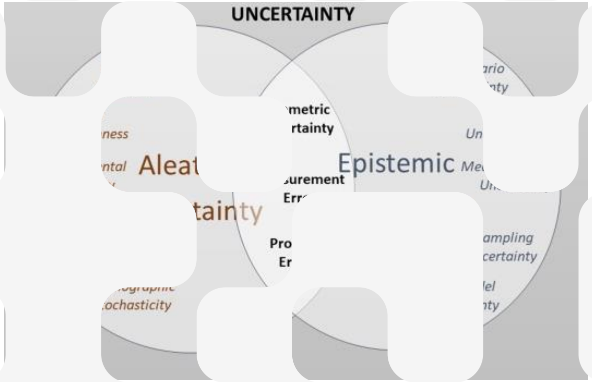
Agenda

I. RECAP ON CNNs

II. SEMANTIC SEGMENTATION

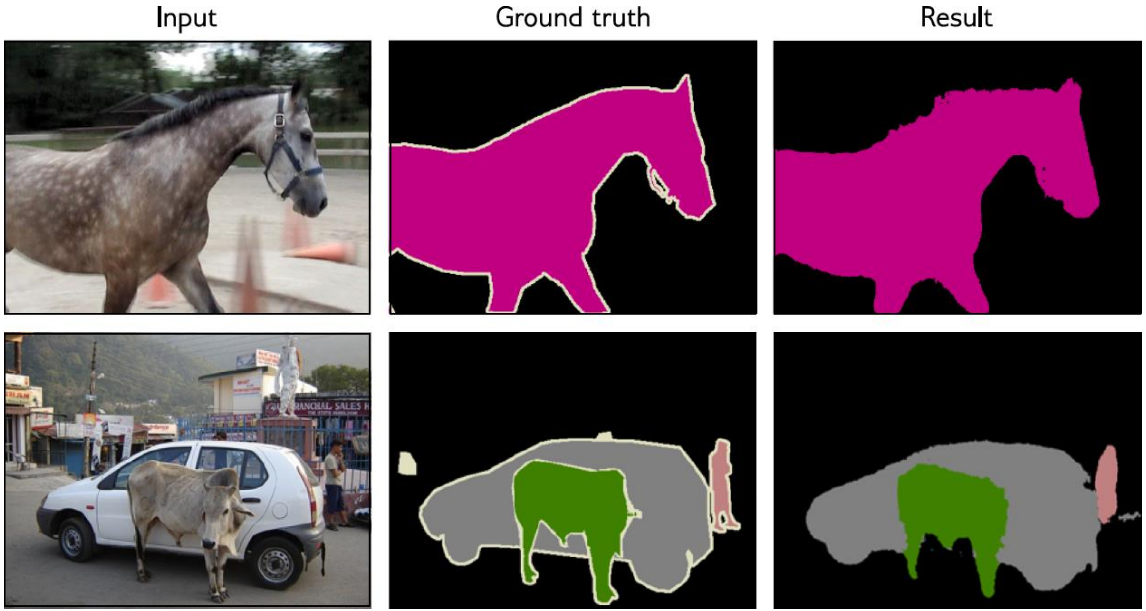
III. OBJECT DETECTION

IV. TIPS AND TRICKS IN ML



All models are wrong, but ~~some~~ models that know when they are wrong, are useful
/George Box + unknown researcher from the Google AI Brain Team/

Semantic segmentation



S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

Semantic segmentation: key concepts

Semantic segmentation involves labeling each pixel in an image with a corresponding class label from a predefined set of classes.

1. Pixel-Level Classification:

Each pixel in the image is assigned a class label. For example, in an image of a street scene, pixels might be labeled as "road," "car," "pedestrian," "sidewalk," etc.

2. Applications:

Autonomous Driving: Identifying road lanes, pedestrians, vehicles, and other objects.

Medical Imaging: Segmenting different tissues, organs, and abnormalities.

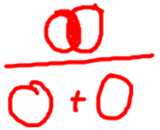
Robotics: Understanding and navigating environments.

Augmented Reality: Overlaying virtual objects accurately on real-world scenes.

Surveillance: Tracking objects and understanding their movements.

3. Evaluation metric:

Intersection over Union (IoU) also known as Jaccard Index, IoU measures the overlap between the predicted segmentation and the ground truth. It is calculated as: *IoU=Area of Overlap / Area of Union.*

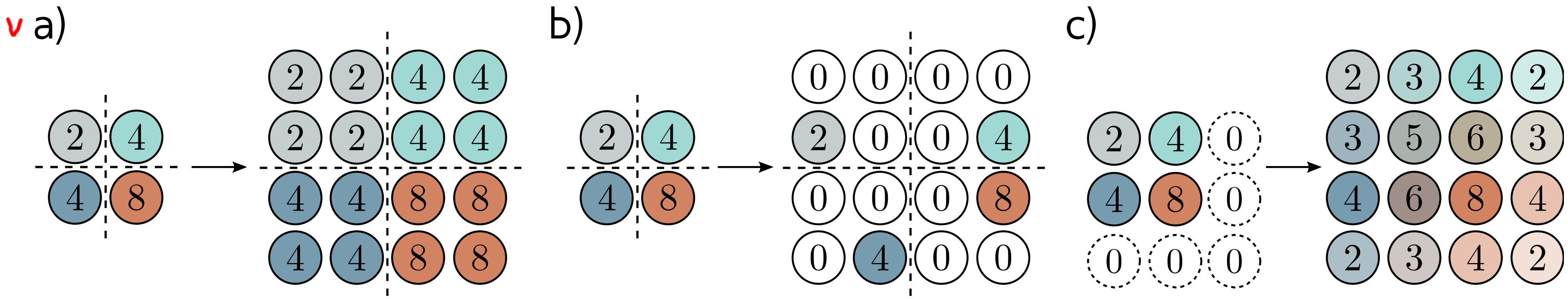


Some details of a segmentation model: upsampling, transposed convolutions

Duplicate

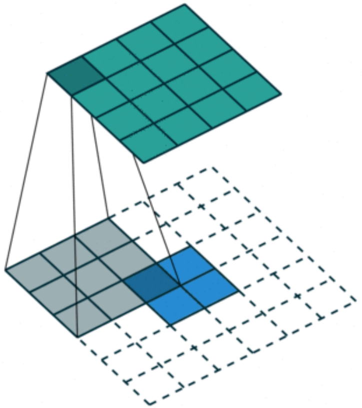
Max-upsampling

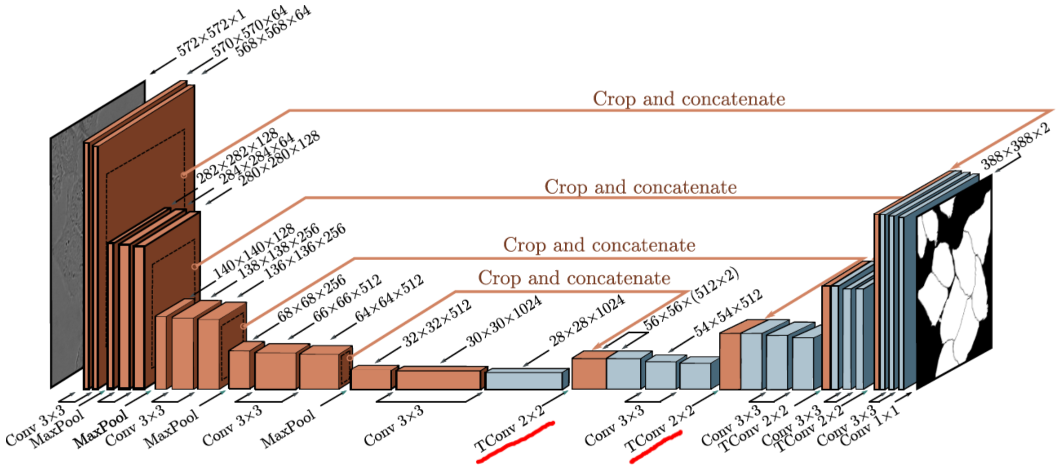
Bilinear interpolation



S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

Some details of a segmentation model: upsampling, transposed convolutions





[S. J. Prince. Understanding Deep Learning. MIT Press, 2023.]

ML-2024
by Alexei Kornaev

Loss functions

Commonly used loss functions include:

•**Cross-Entropy Loss**: For binary and multi-class segmentation.

•**Dice Loss**: For emphasizing overlap between predicted and ground truth segmentation.

•**Tversky Loss**: For balancing false positives and false negatives.

•**Focal Loss**: For addressing class imbalance.

•**Lovász-Softmax Loss**: For optimizing IoU directly.

•**Combo Loss**: For combining multiple loss functions.

Each loss function has its own advantages and is suitable for different scenarios. Experimenting with different loss functions and tuning hyperparameters can lead to better performance in semantic segmentation tasks.

[Generated by DeepSeek-V2.5]

1. **Cross-Entropy Loss**

- **Binary Cross-Entropy Loss**:
  - Used for binary segmentation tasks (e.g., foreground vs. background).
  - Formula:

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i\log(h_i) + (1-y_i)\log(1-h_i)\right]$$

    where y_i is the ground truth label, h_i is the predicted probability, and N is the number of pixels.

- **Categorical Cross-Entropy Loss**:
  - Used for multi-class segmentation tasks.
  - Formula:

$$L_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(h_{ic})$$

    where y_{ic} is the ground truth label for pixel i and class c, h_{ic} is the predicted probability for class c, N is the number of pixels, and C is the number of classes.

2. **Dice Loss (F1 Score Loss)**

- **Dice Coefficient**:
  - Measures the overlap between the predicted segmentation and the ground truth.
  - Formula:

$$\mathrm{Dice} = \frac{2\cdot|X\cap Y|}{|X|+|Y|}$$

    where X is the predicted segmentation and Y is the ground truth.

- **Dice Loss**:
  - The Dice loss is the complement of the Dice coefficient.
  - Formula:

$$L_{\mathrm{Dice}} = 1 - \mathrm{Dice}$$

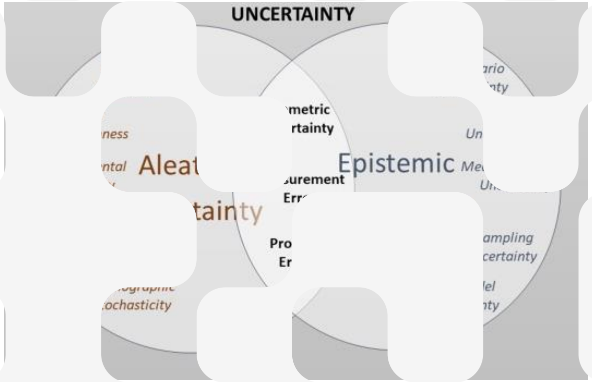
Agenda

I. RECAP ON CNNs

II. SEMANTIC SEGMENTATION

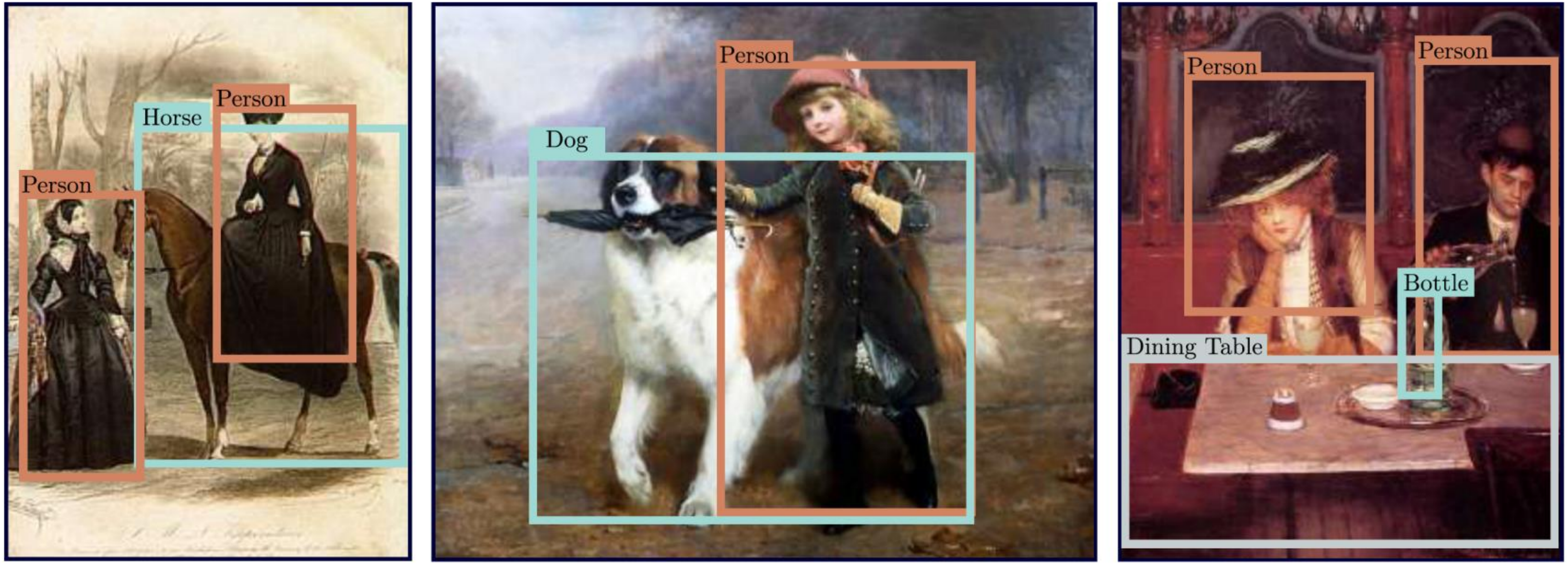
III. OBJECT DETECTION

IV. TIPS AND TRICKS IN ML



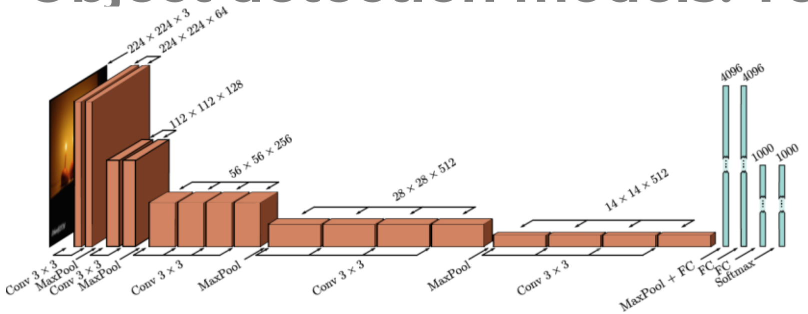
All models are wrong, but ~~some~~ models that know when they are wrong, are useful
/George Box + unknown researcher from the Google AI Brain Team/

Object detection models

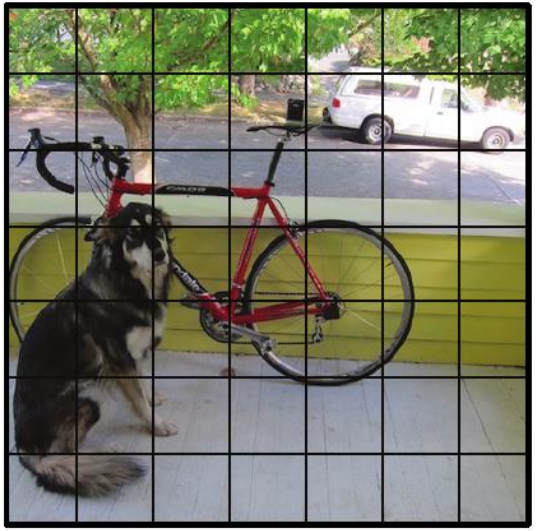


[S. J. Prince. Understanding Deep Learning. MIT Press, 2023.]

Object detection models: You Only Look Once (YOLO)



- Network similar to VGG (448x448 input)
- 7×7 grid of locations
- Predict class at each location
- Predict 2 bounding boxes at each location
  - Five parameters –x,y, height, width, and confidence
- Momentum, weight decay, dropout, and data augmentation
- Heuristic at the end to threshold and decide final boxes

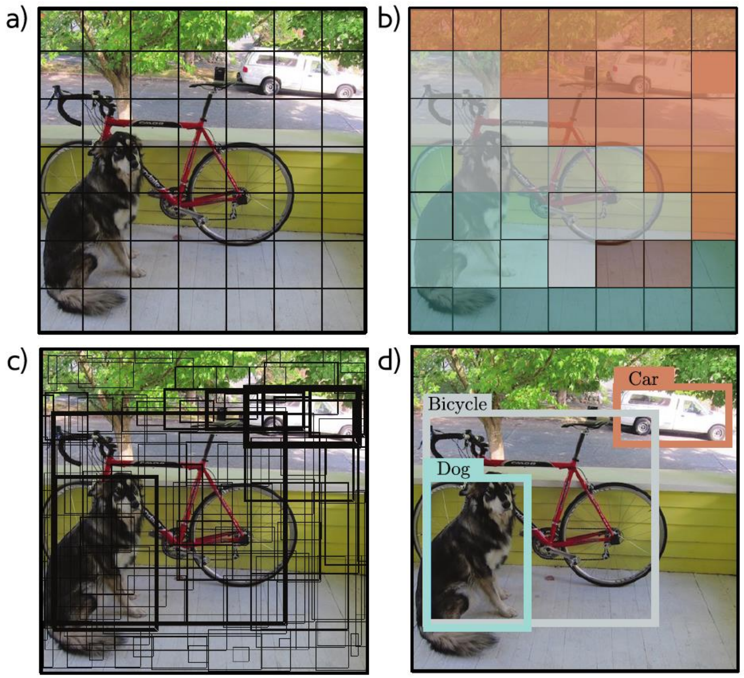


[S. J. Prince. Understanding Deep Learning. MIT Press, 2023.]

Object detection models: You Only Look Once (YOLO)

- Network similar to VGG (448x448 input)
- 7×7 grid of locations
- Predict class at each location
- Predict 2 bounding boxes at each location
  - Five parameters –x,y, height, width, and confidence
- Momentum, weight decay, dropout, and data augmentation
- Heuristic at the end to threshold and decide final boxes

x,y w h



S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

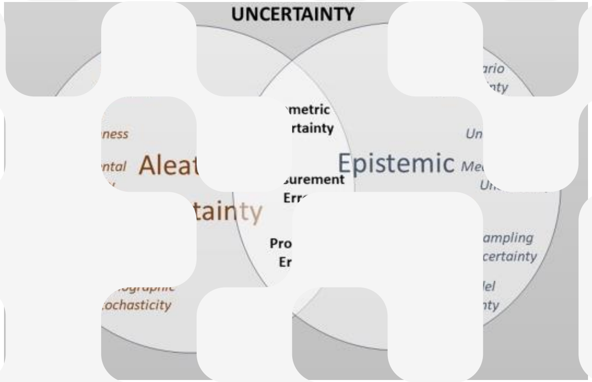
Agenda

I. RECAP ON CNNs

II. SEMANTIC SEGMENTATION

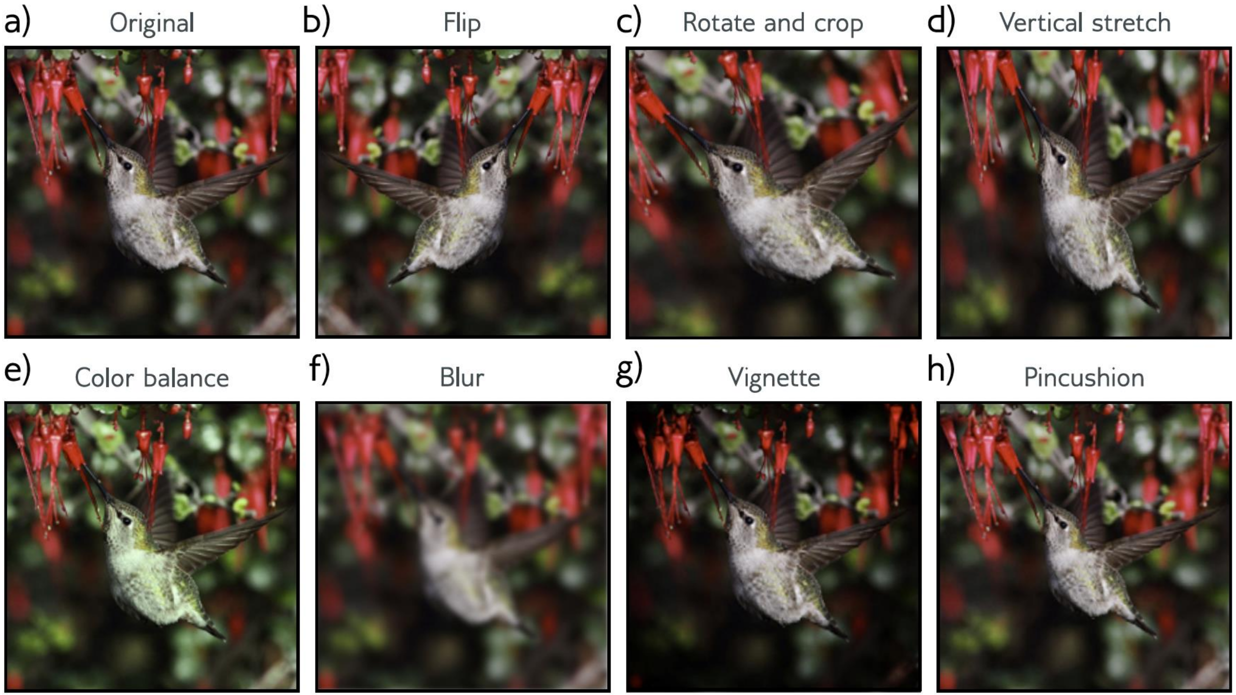
III. OBJECT DETECTION

IV. TIPS AND TRICKS IN ML



All models are wrong, but ~~some~~ models that know when they are wrong, are useful
/George Box + unknown researcher from the Google AI Brain Team/

Data augmentation

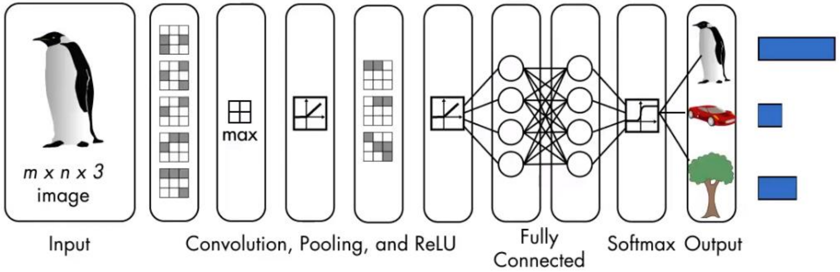


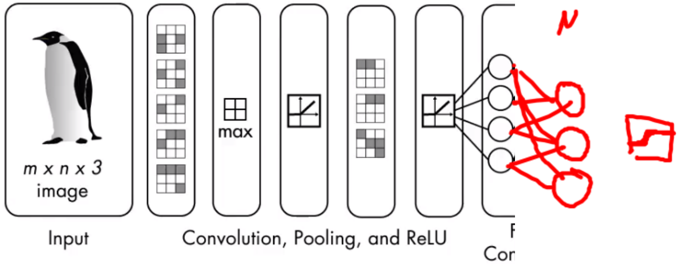
[S. J. Prince. Understanding Deep Learning. MIT Press, 2023.]

Transfer learning

1. Random init
2. Fine-tuning
3. Transfer learning

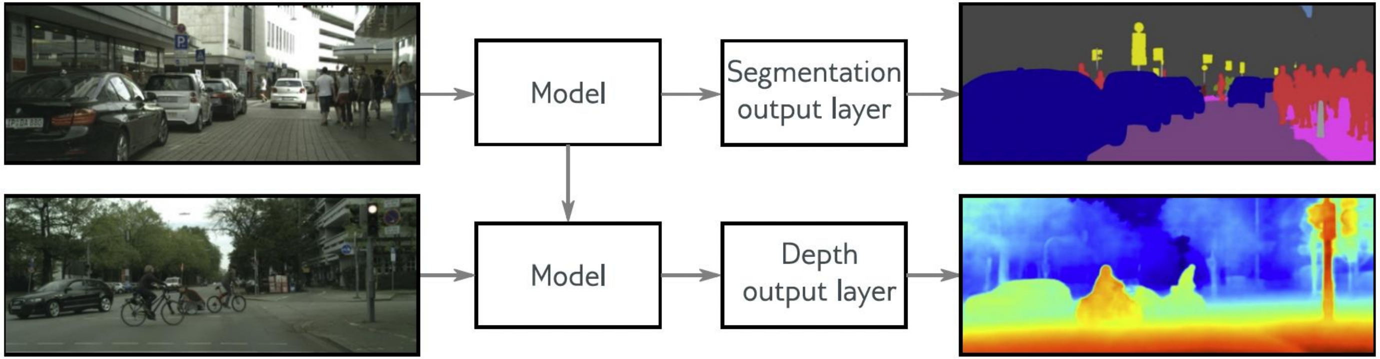
$\alpha_1 < \alpha_2 < \alpha_3 \ldots$





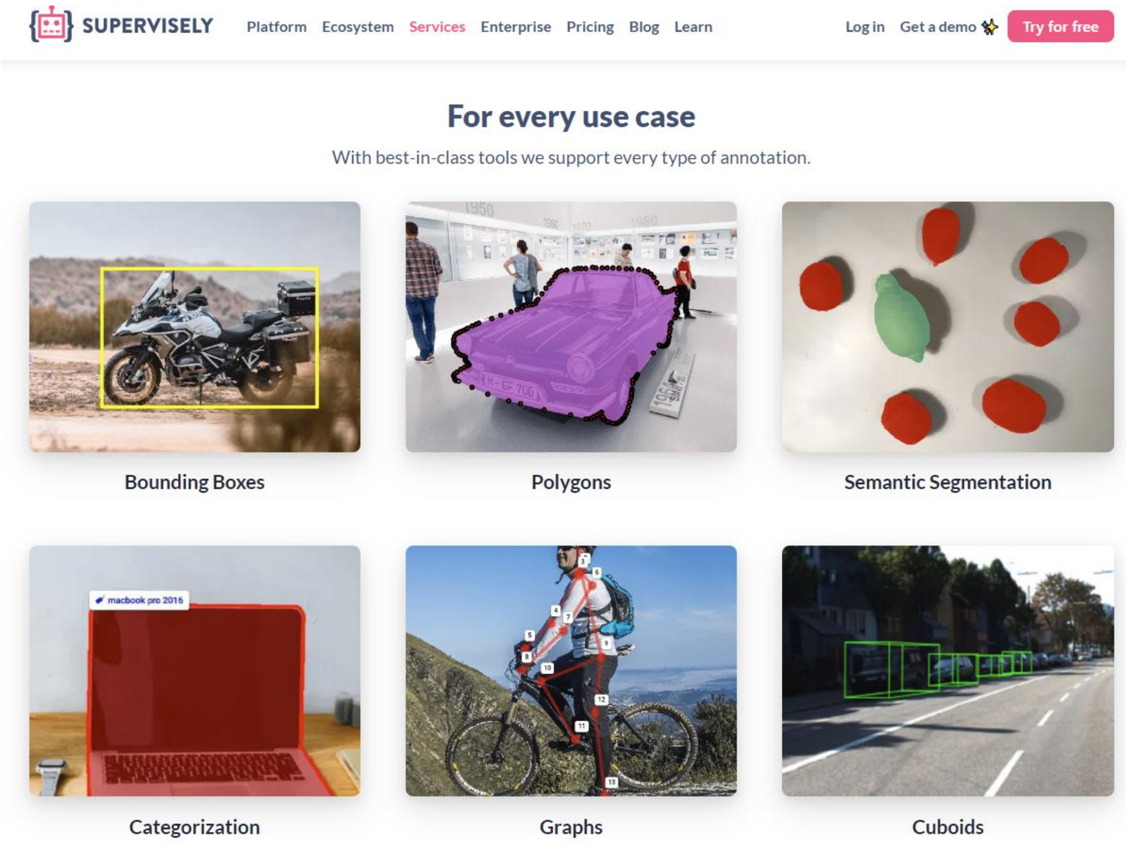
S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

Transfer learning



S. J. Prince. Understanding Deep Learning. MIT Press, 2023.

Labeling



[Supervise.ly]()

What to do if your model is prone to overfitting?

1. Regularization (L1, L2)
2. Architectures: CNNs, ResNets
3. Transfer learning vs random initialization
4. Data augmentation (train-time a., test-time a.)
5. Dropout (train-time DO, test-time DO)
6. Ensembling

Thank you for your attention!

a.kornaev@innopolis.ru, @avkornaev