Object Detection in Snorkel
Data Programming – Labelling functions

• Express knowledge as labelling functions
  • Can have unknown accuracies and correlations
  • Assign a class label or abstain

```python
def LF_wife_in_sentence(c):
    """A simple example of a labeling function""
    return 1 if 'wife' in c.get_parent().words else 0
```
Generative Model

- Probabilistic Model
  - Optimized by minimizing the negative log marginal likelihood given the observed label matrix $\Lambda$
    
    $$p_w(\Lambda, Y) = Z_w^{-1} \exp \left( \sum_{i=1}^{m} w^T \phi_i(\Lambda, y_i) \right)$$

- Generate probabilistic training labels
    
    $$\tilde{Y} = p_{\hat{w}}(Y | \Lambda)$$
Object Detection
Object Detection

• Localize and classify objects in an image
• Multiclass & Variable number of labels
Architectures

• Region-based Convolutional Network
  • More accurate, Slow

• You Only Look Once (YOLO)
  • YOLO (2016), YOLOv2 (2016), YOLOv3 (2018)
  • Less accurate, Real-time
Object Detection in Snorkel
Approach

• Regions as candidates
• Models as labelling functions
• Combining labels from different detection models
• Extension with image classification models
• Augmenting the dataset with new images
Regions as candidates

- Each detected region is treated as a candidate
Models as labelling functions

• Treat each trained model as a labelling function

```python
def LF_resnet(im):
    return resnet_model.predict(im)

def LF_inception(im):
    return inception_model.predict(im)
```
Combining labels from different models

- Object detection models
  - Combine regions based on IoU (intersection over union) measure
  - Abstain for non-detected regions

![Diagram showing IoU calculation](image)
Extension with image classification models

- Apply simple image classification models to the candidate regions
Augmenting the existing dataset

• Train a generative model based on the labels
• New training samples from new images using the generative model
Evaluation

• Retrain existing machine learning models with augmented dataset
• Effects of extra data
• Effects of using probabilistic labels
• Manual investigation of random samples