

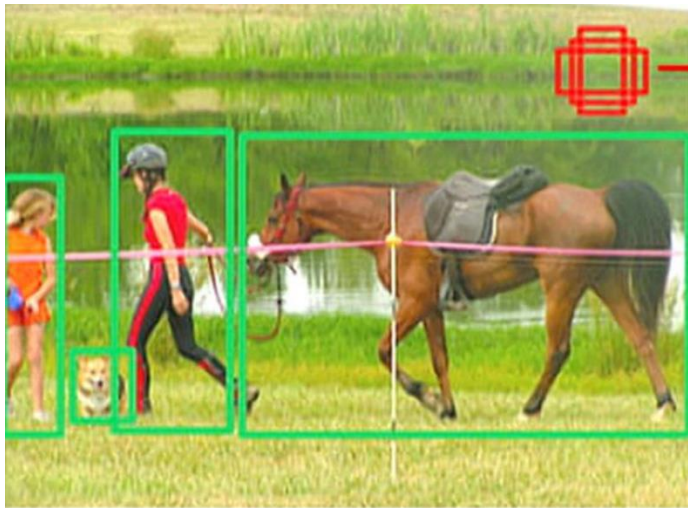
Losses Part 2: Advanced Loss Functions

Contents

- Focal Loss – Class Imbalance
- ArcFace – Face Recognition
- CTC – Sequence Alignment
- Chamfer and Dice – 3D + Segmentation
- KL Divergence – Distribution Matching
- RLHF – LLM Alignment

Focal Loss : Solving Extreme Class Imbalance

- 100 predicted boxes → Only 7 contain objects
- 93 boxes are background (negative examples)
- Ratio: 1:13 (can be 1:100 in practice!)



Many negative examples, no useful signal

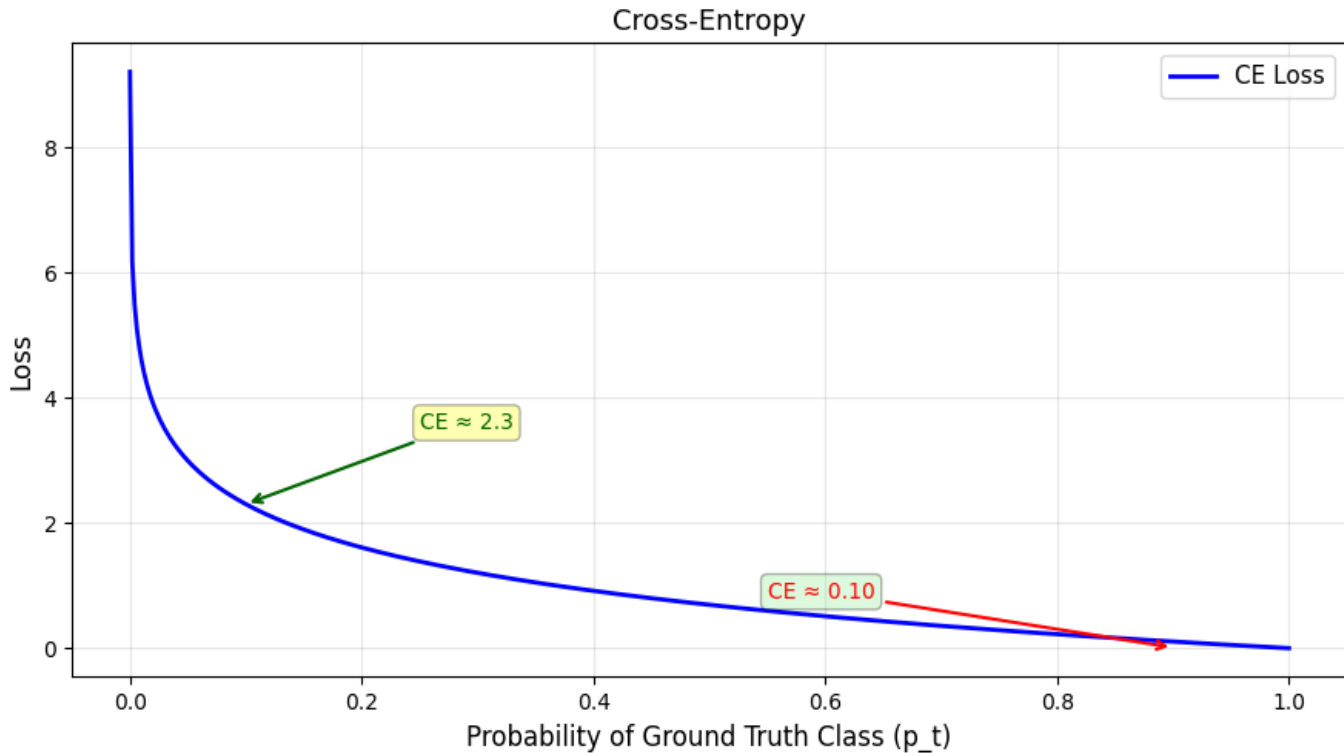
Few positive examples, rich information

Why Cross-Entropy Fails?

→ $100,000 \text{ easy} \times 0.1 = 10,000 \text{ total loss}$

- Easy examples dominate!
- 98% of gradient signal wasted.

→ $100 \text{ hard} \times 2.3 = 230 \text{ total loss}$

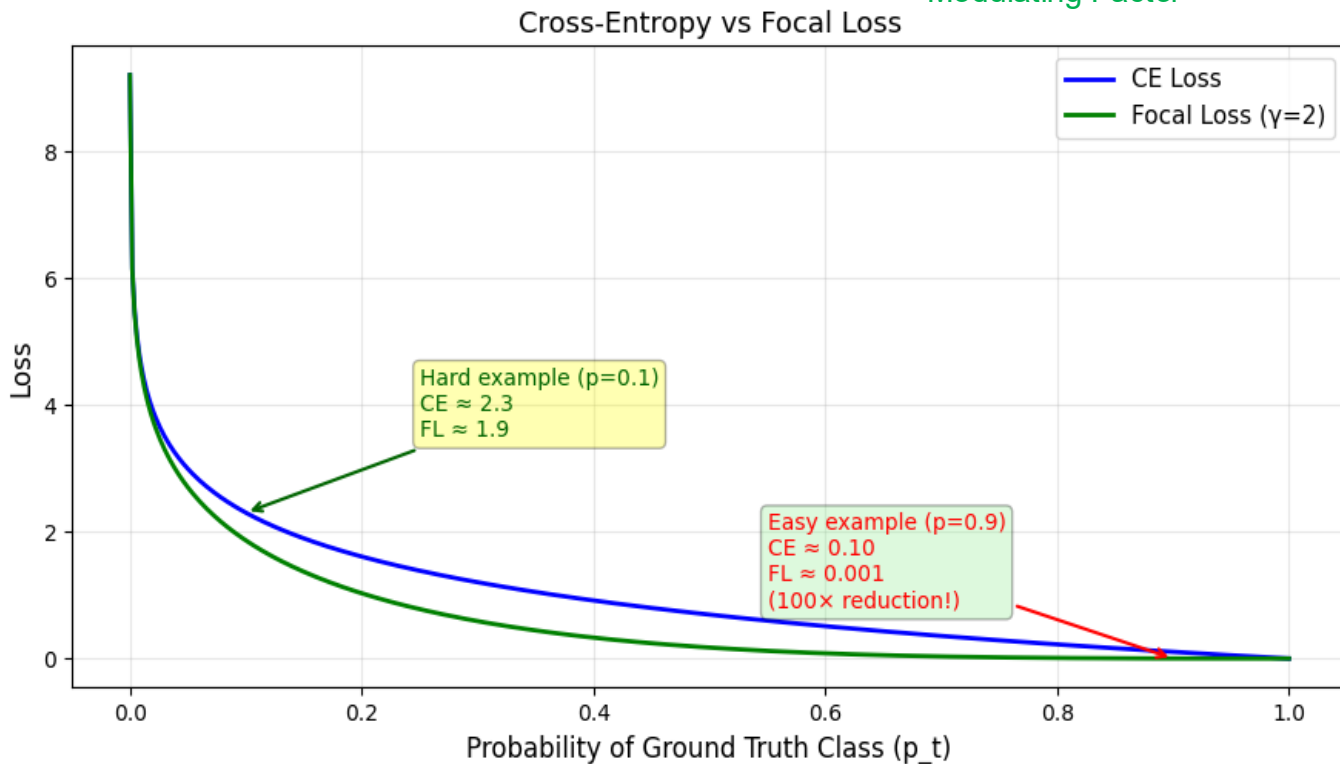


Focal Loss : The Solution

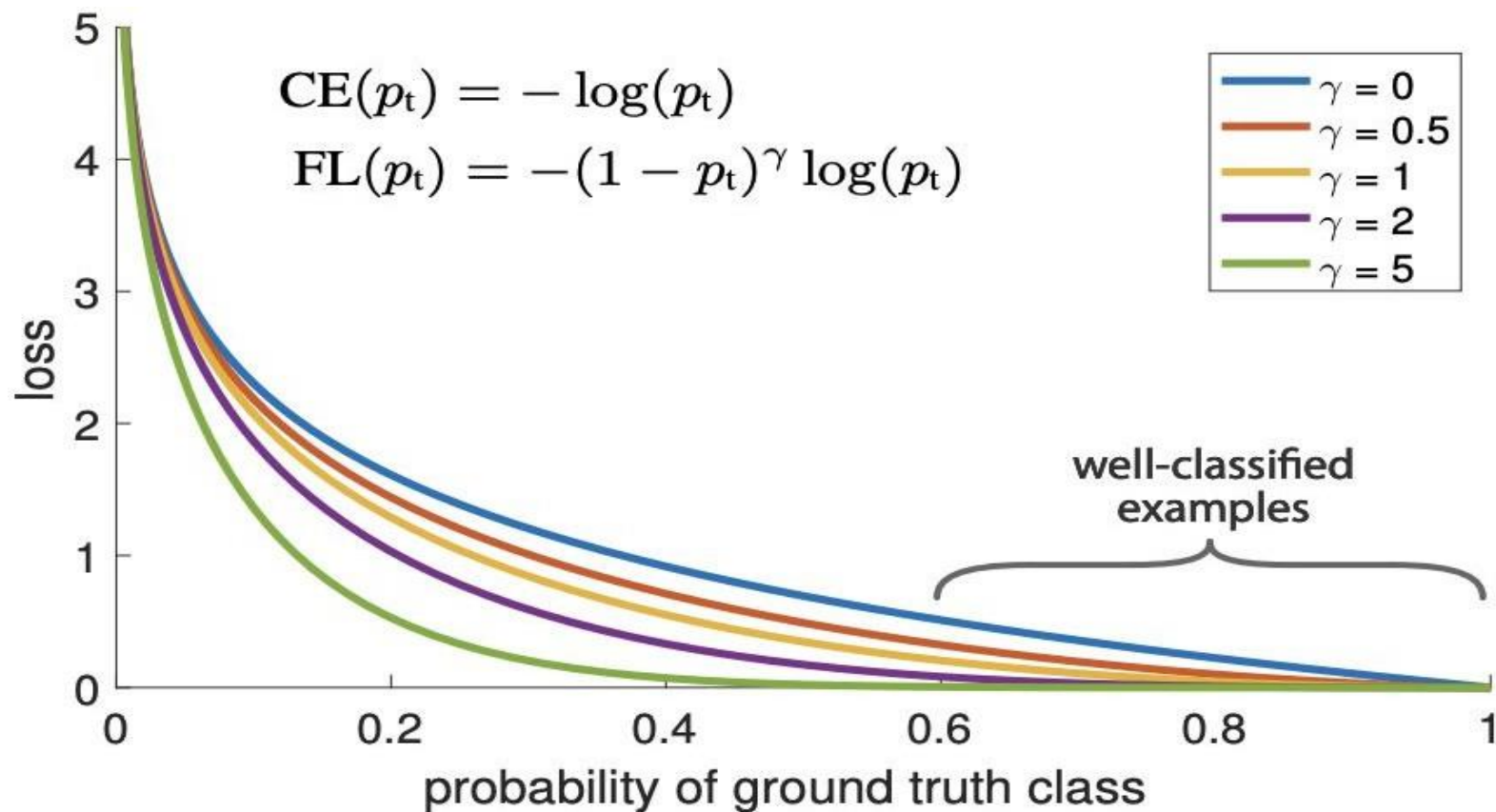
$$\text{CE}(p_t) = -\log(p_t)$$

$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

Modulating Factor



Tuning γ : How Aggressive Should We Be?



ArcFace Loss : Face Recognition Without Retraining

11-785

Introduction to
Deep Learning

Training: 10,000 people in dataset

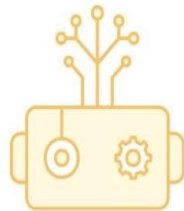
Testing: New person walks in tomorrow

Problem: Traditional classifier has 10,000 outputs

Where does person #10,001 go?

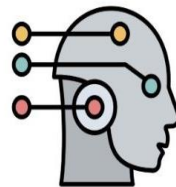
Solution: Learn embeddings

- Similar faces \rightarrow close vectors
- Different faces \rightarrow far vectors



How do we make it tell two faces apart?

In comes ArcFace Loss!



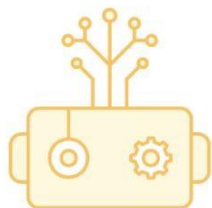
We train an AI to detect human faces

As part of HW2P2, we train a model to classify if the image has a human face or not.

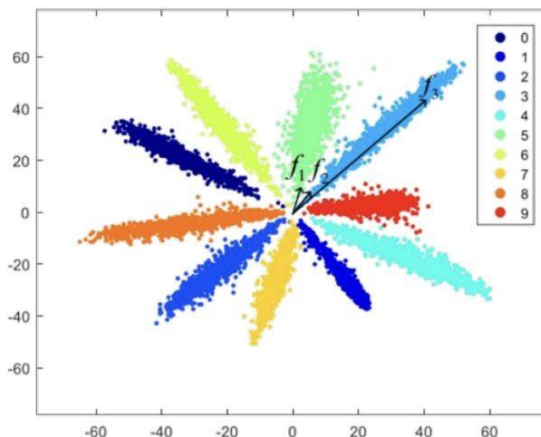
Problem : Naïve Approaches Fail!

11-785

Introduction to
Deep Learning



Model learned to
represent faces



Each color is a unique face



Cosine Similarity?

Won't work well

Due to magnitude
F1, F2 more similar
than F2, F3

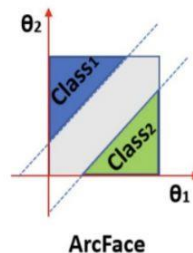
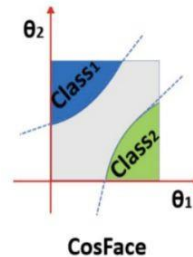
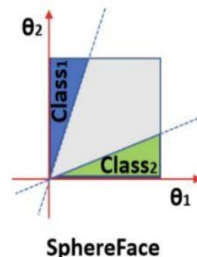
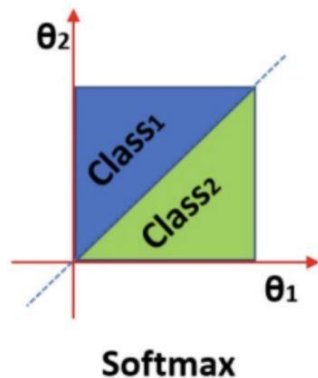
Need angular separation, not just Euclidean distance

ArcFace: Adding Angular Margin

11-785

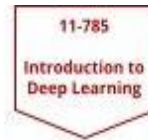
Introduction to
Deep Learning

Softmax learns to
separate

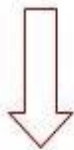


We need to add
margin in between
the separated
classes!

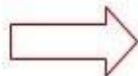
CTC Loss : Sequence Alignment Without Labels



Multimodal Learning
is all the rage?



Let's solve this
video-to-text task

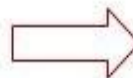


Input



Output

2025



We have frames that
represent a digit

Predict digit for each frame,
Merge together duplicates

We Solve It!

CTC: The Blank Token Trick

11-785

Introduction to
Deep Learning

Multimodal Learning
is all the rage?



Let's solve this
video-to-text task



Predict digit for each frame

2 2 2 2 2 0 0 0 2 2 2 2 5 5 5 5



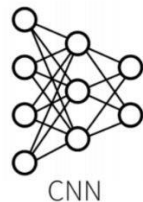
Merge duplicates

2 0 2 5

Chamfer Loss : Comparing Point Clouds



Input Image



CNN



Predicted points

↔
Loss



Sampled GT points

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2$$

Can handle different number of points in (prediction, GT)

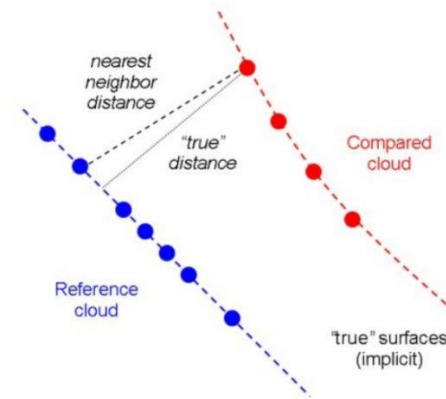


Image courtesy: cloudcompare

[Link](#) to documentation

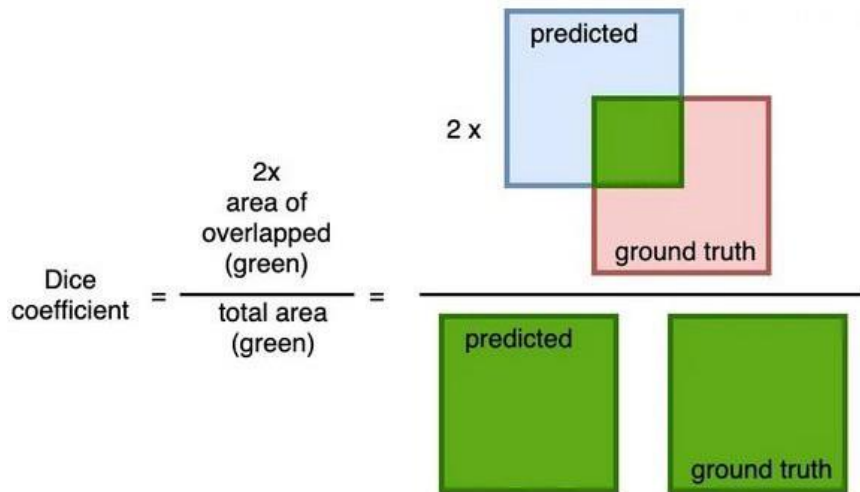
Dice Loss : Segmentation with Extreme Imbalance

Medical imaging: Brain tumor detection

Tumor pixels: 100 (0.1%)

Background: 99,900 (99.9%)

- Cross-Entropy:
Predicts "all background" → 99.9% accuracy!
Completely useless!
- Dice: Focuses on overlap, not pixel counts



$$\mathcal{L}_{\text{Dice}} = 1 - \text{DiceCoef}$$

KL Divergence : Measuring Distribution Difference

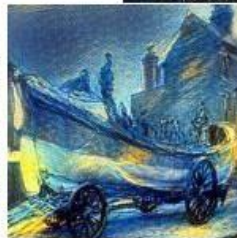
Perceptual, KL-Divergence, ELBO

Question: How different are two probability distributions?

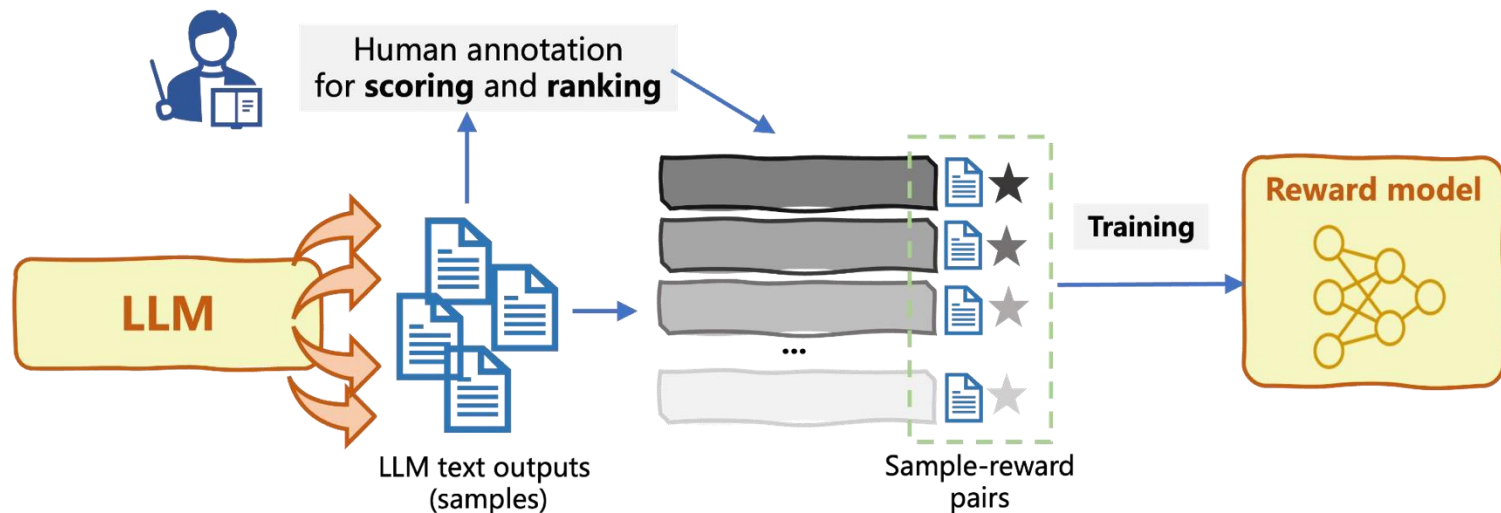
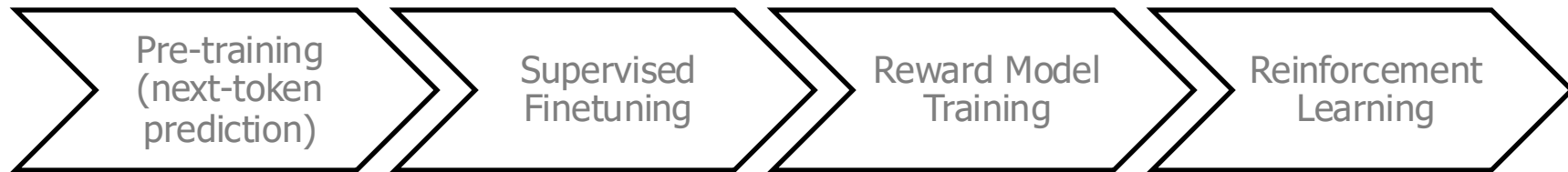
$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Information lost when using Q to approximate P

Style
The Starry Night,
Vincent van Gogh,
1889



RLHF : How ChatGPT Learned to Be Helpful



Remember!

Loss function = Model's compass → So, choose wisely!

Homework Implementations –

HW2: Face Recognition with ArcFace

→ See the accuracy improvement!

HW3: Handwriting Recognition with CTC

→ Learn sequence alignment automatically

Final Project: Mix & Match

→ Object detection? Focal Loss

→ Segmentation? Dice Loss

→ Generative? KL Divergence