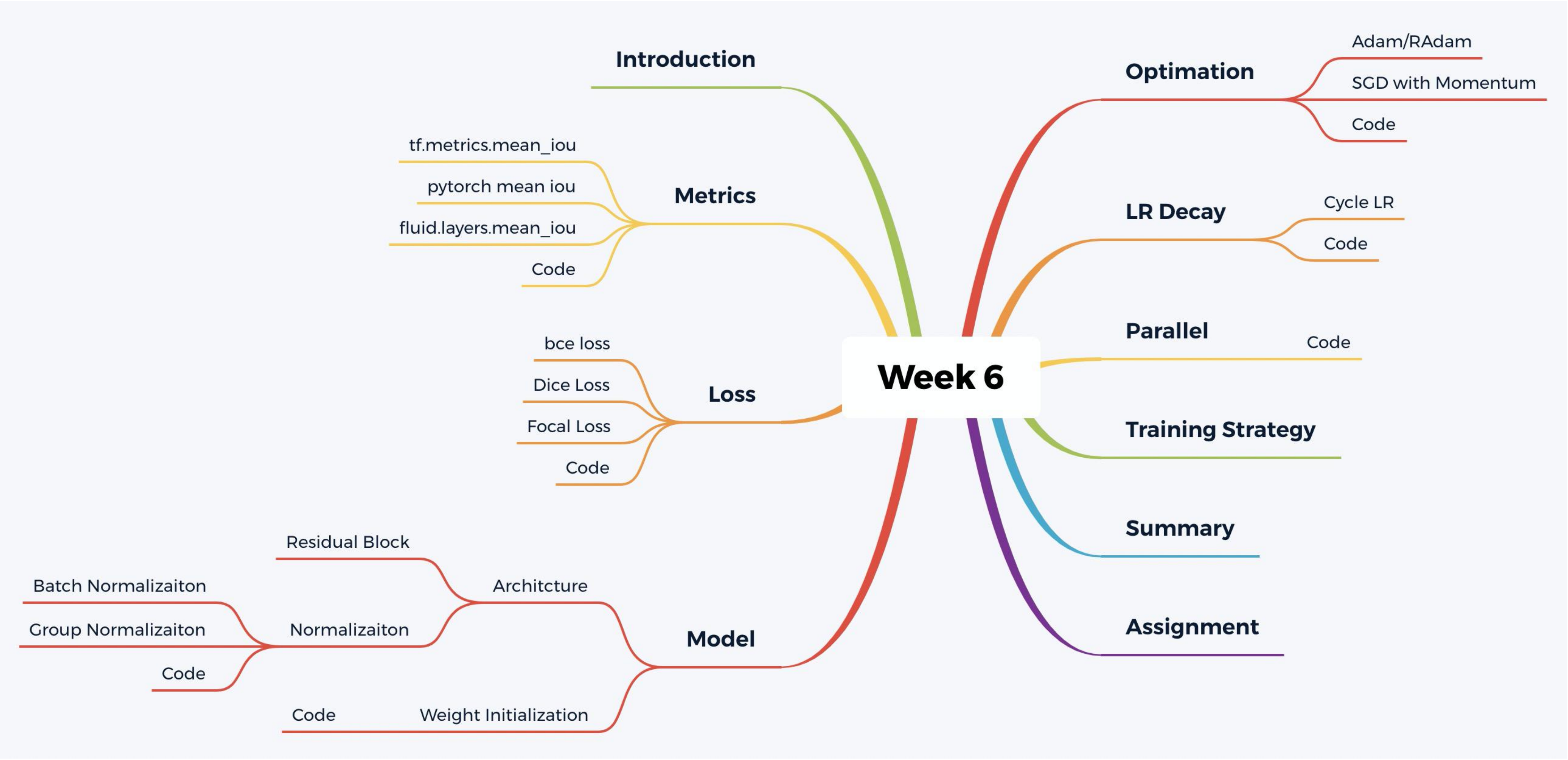
Lane Segmentation Week 6

HCT CV Course

主要内容


Week 6
Introduction
Metrics
tf.metrics.mean_iou
pytorch mean iou
fluid.layers.mean_iou
Code
Loss
bce loss
Dice Loss
Focal Loss
Code
Model
Architcture
Residual Block
Normalizaiton
Batch Normalizaiton
Group Normalizaiton
Code
Weight Initialization
Code
Optimation
Adam/RAdam
SGD with Momentum
Code
LR Decay
Cycle LR
Code
Parallel
Code
Training Strategy
Summary
Assignment

学习目标

- 理解Semantic Segmentation Metrics
- 理解Semantic Segmentation Loss
- 理解优化算法
- 理解调整学习率的算法
- 理解权重初始化的方法
- 掌握单卡/多卡训练方法
- 掌握分阶段训练的策略

training strategy

- Collect more data
- Train algorithm longer
- Try Adam instead of gradient descent
- Add regularizations
- Network architecture
- ……

training strategy

- One of the challenges with building machine learning systems is that there's so many things you could try, so many things you could change.
- hyperparameters

The No Free Lunch Theorem

- Learning theory claims that a machine learning algorithm can generalize well froma finite training set of examples. This seems to contradict some basic principles of logic.

The No Free Lunch Theorem

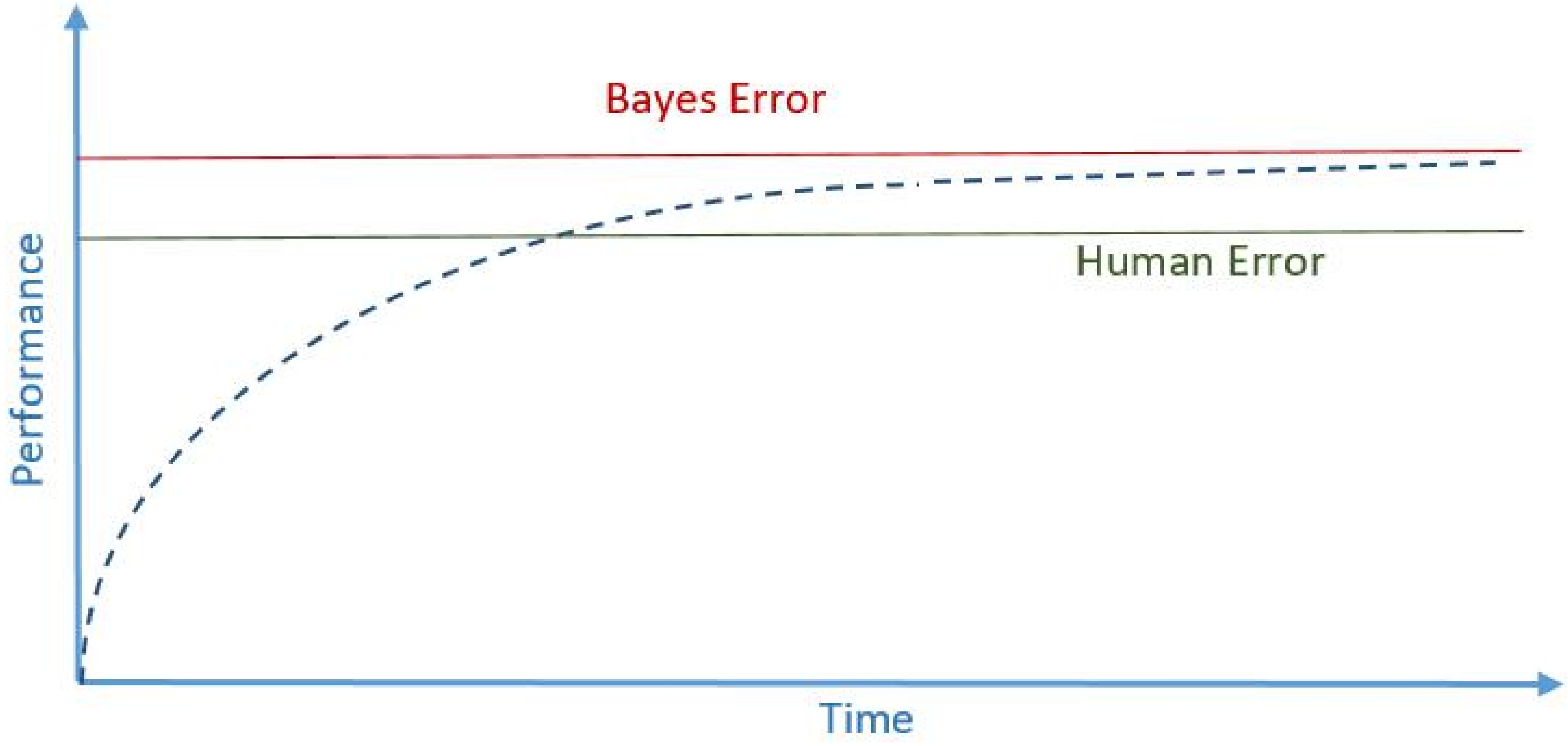
- no machine learning algorithm is universally any better than any other

The No Free Lunch Theorem

- If we make assumptions about the kinds of probabilitydistributions we encounter in real-world applications, then we can design learningalgorithms that perform well on these distributions.

Watch your data !!!!!!

Bayes Error



human-level performance

- define your key priorities
- based on observations of performances and dataset

Bias

- human-level performance
- Underfitting

Variance

- Overfitting

Training

- 观察数据
- 确定Baseline
- 对Baseline模型进行优化
- ablation study

Baseline

- no data augmentation
- no big model
- low resolution
- little tricks

Baseline

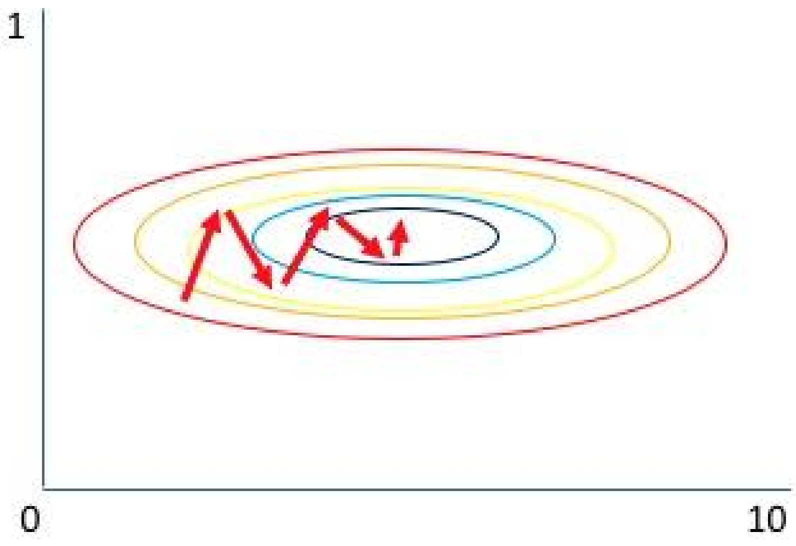
- https://github.com/gujingxiao/Lane-Segmentation-Solution-For-BaiduAI-Autonomous-Driving-Competition/blob/master/train.py

Training

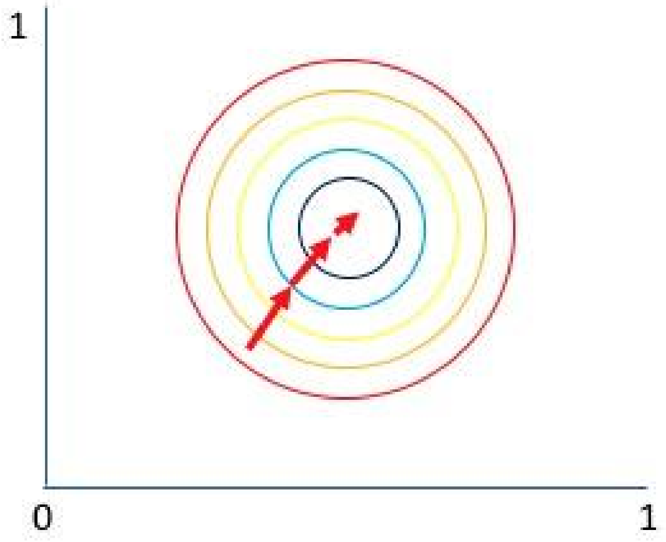
- Data Generator
- Metrics
- Loss
- Model
- Optimazition

Orthogonalization

Normalization



Gradient of larger parameter
dominates the update



Both parameters can be
updated in equal proportions

Note

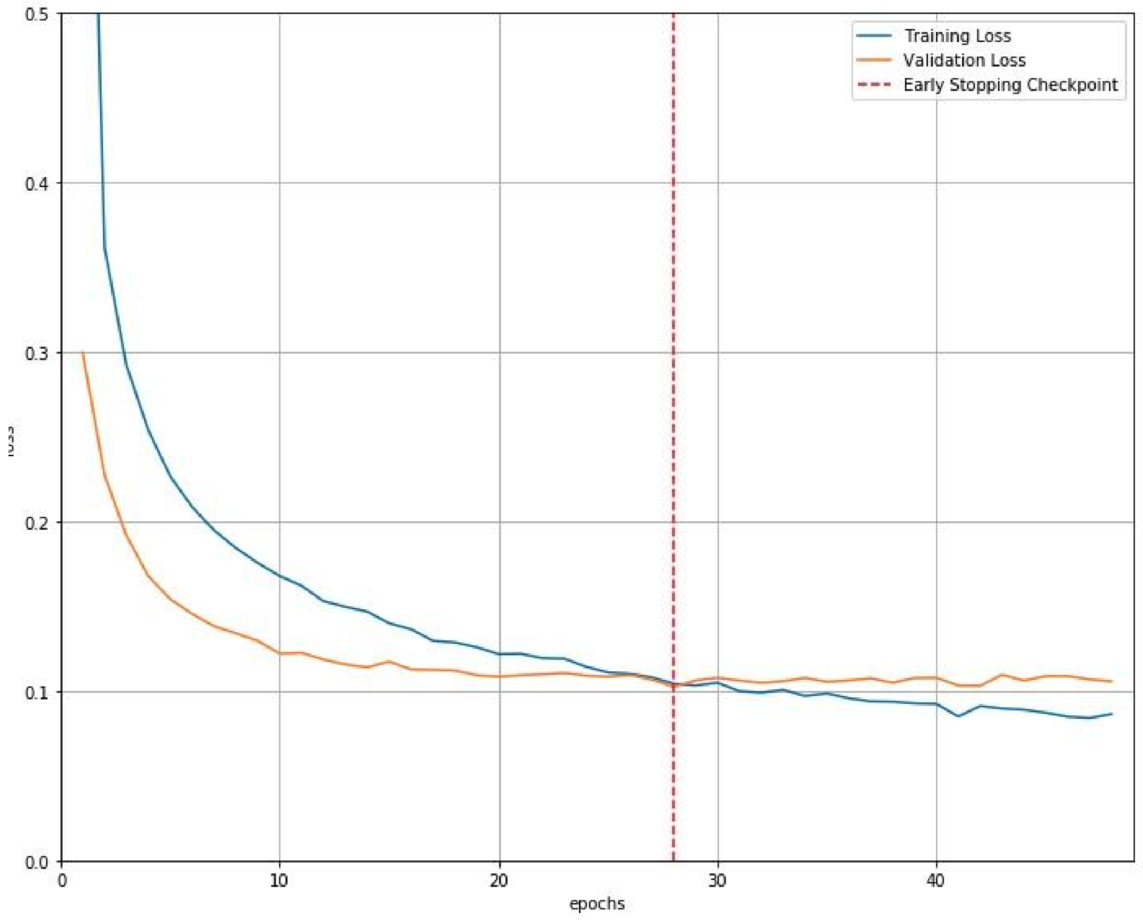
number of training epochs

- Too many epochs can lead to overfitting of the training dataset, whereas too few may result in an underfit model. Early stopping is a method that allows you to specify an arbitrary large number of training epochs and stop training once the model performance stops improving on a hold out validation dataset.

Early-Stop

```
callback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=3)
# This callback will stop the training when there is no improvement in
# the validation loss for three consecutive epochs.
model.fit(data, labels, epochs=100, callbacks=[callback],
    validation_data=(val_data, val_labels))
```

Early-Stop



Training Loss
Validation Loss
Early Stopping Checkpoint
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
epochs

save model

- early stop

单卡/多卡

Model

- MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation
- https://arxiv.org/abs/1902.04049

DeepLab v3+

- BN/GN/LRN

Metrics

- CityScape

  https://www.cityscapes-dataset.com/benchmarks/

TP/FP/FN/TN

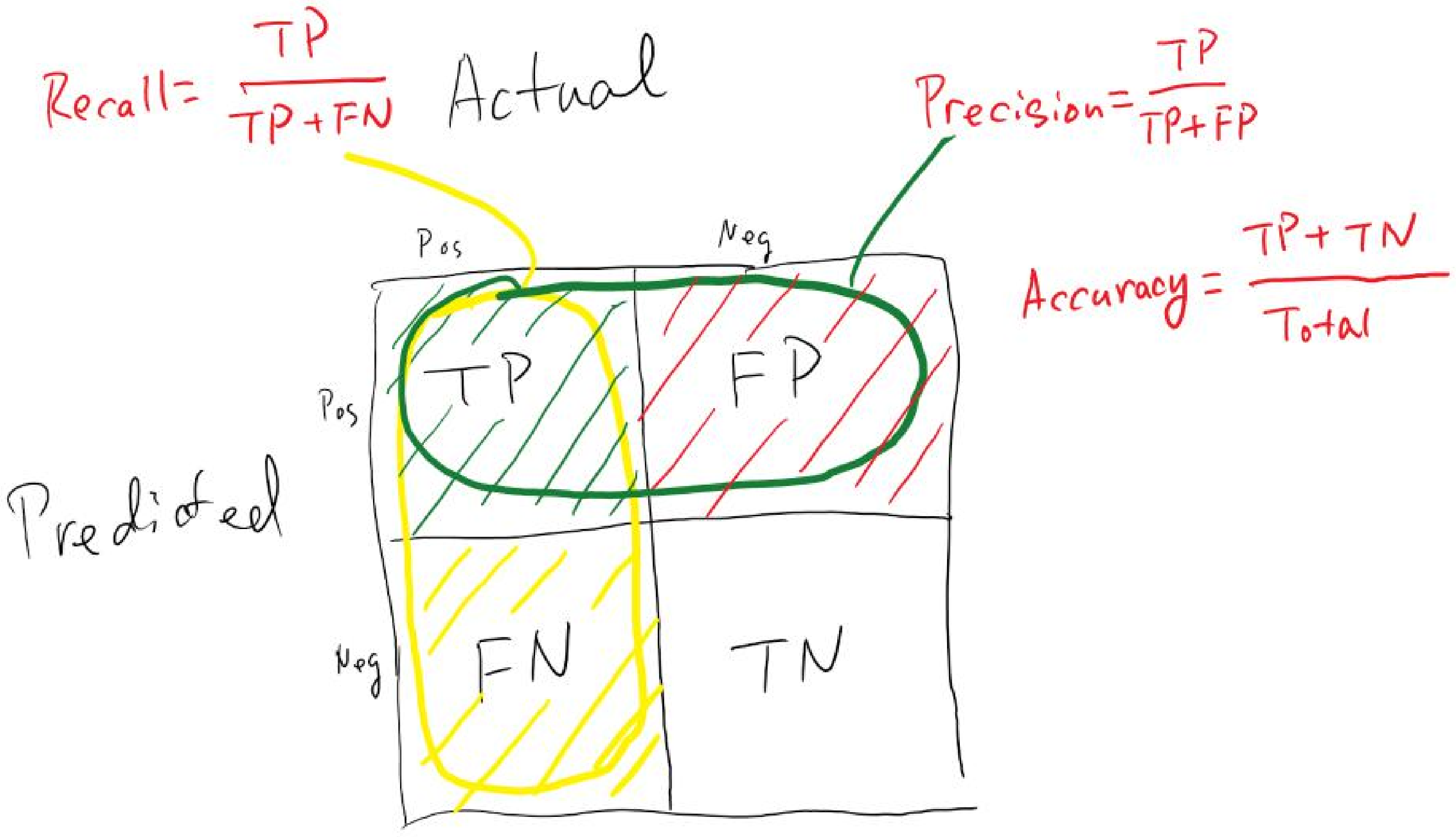
我们可以使用一个 2x2 混淆矩阵来总结我们的"狼预测"模型，该矩阵描述了所有可能出现的结果（共四种）：

真正例 (TP)： • 真实情况：受到狼的威胁。 • 牧童说："狼来了。" • 结果：牧童是个英雄。	**假正例 (FP)：** • 真实情况：没受到狼的威胁。 • 牧童说："狼来了。" • 结果：村民们因牧童吵醒他们而感到非常生气。
假负例 (FN)： • 真实情况：受到狼的威胁。 • 牧童说："没有狼"。 • 结果：狼吃掉了所有的羊。	**真负例 (TN)：** • 真实情况：没受到狼的威胁。 • 牧童说："没有狼"。 • 结果：大家都没事。

真正例是指模型将正类别样本正确地预测为正类别。同样，**真负例**是指模型将负类别样本正确地预测为负类别。

假正例是指模型将负类别样本错误地预测为正类别，而**假负例**是指模型将正类别样本错误地预测为负类别。

Precision/Recall



confusion matrix

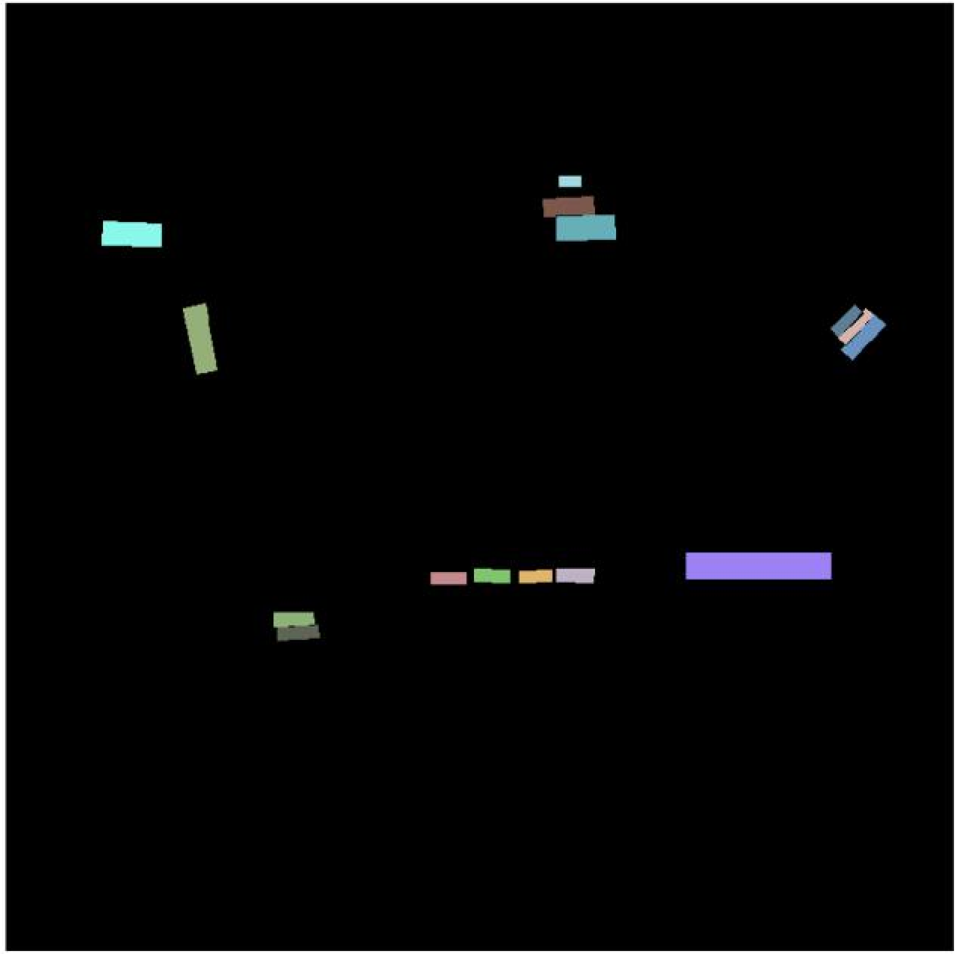
True Class \ Predicted Class	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	923	4	21	8	4	1	5	5	23	6
automobile	5	972	2					1	5	15
bird	26	2	892	30	13	8	17	5	4	3
cat	12	4	32	826	24	48	30	12	5	7
deer	5	1	28	24	898	13	14	14	2	1
dog	7	2	28	111	18	801	13	17		3
frog	5		16	27	3	4	943	1	1	
horse	9	1	14	13	22	17	3	915	2	4
ship	37	10	4	4		1	2	1	931	10
truck	20	39	3	3			2	1	9	923

F-Score

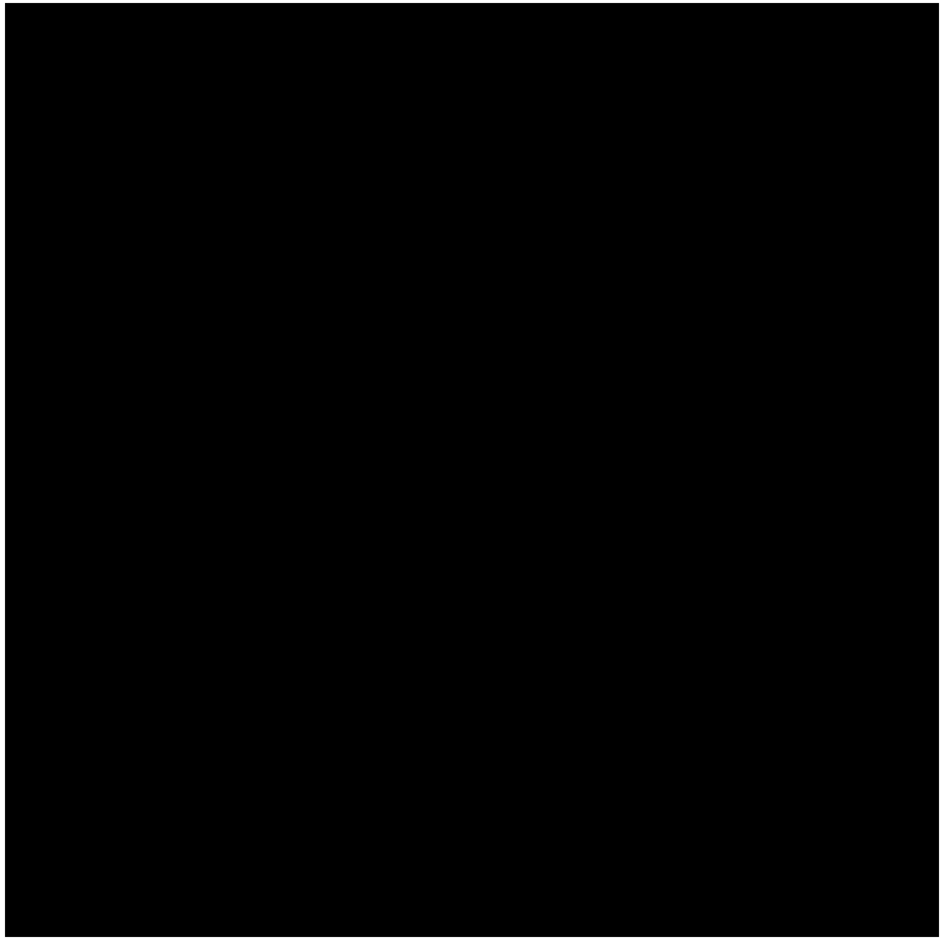
$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 score is the harmonic mean of precision and recall. Values range from 0 (bad) to 1 (good).

Pixel Accuracy



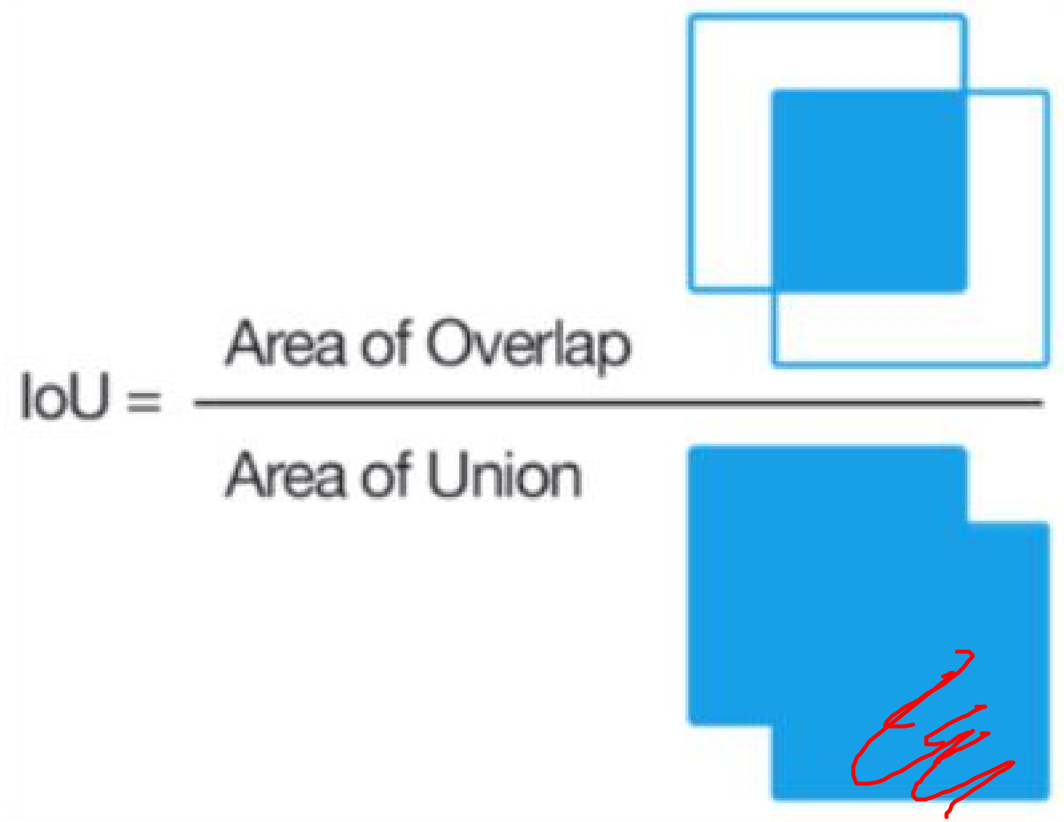
Pixel Accuracy



class imbalance

- Unfortunately, class imbalance is prevalent in many real world data sets, so it can’t be ignored. Therefore, I present to you two alternative metrics that are better at dealing with this issue

Jaccard Index/IoU



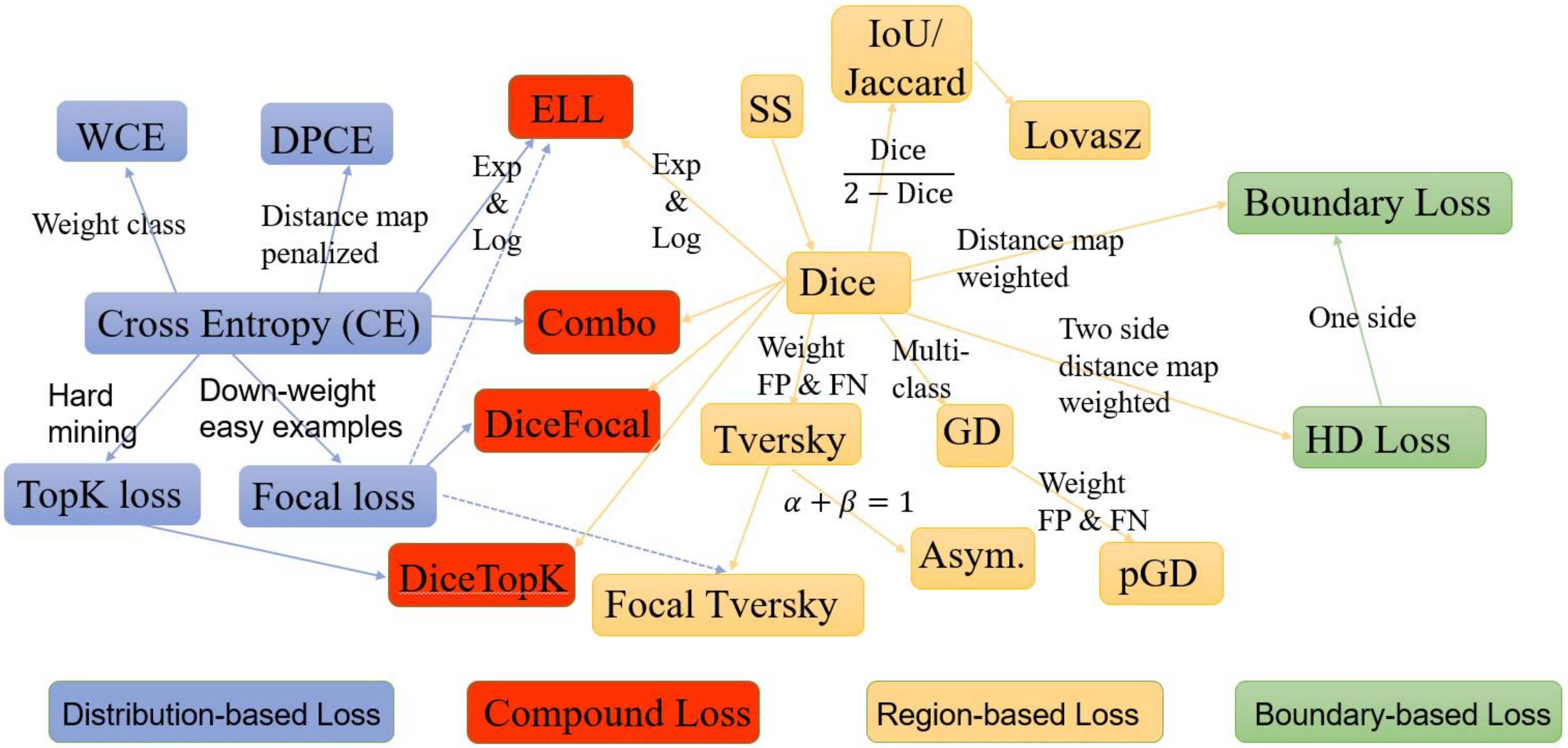
mIoU

- For binary (two classes) or multi-class segmentation, the mean IoU of the image is calculated by taking the IoU of each class and averaging them.

CityScape

- https://www.cityscapes-dataset.com/benchmarks/

Loss



WCE
DPCE
ELL
SS
IoU/ Jaccard
Lovasz
Weight class
Distance map penalized
Exp & Log
Exp & Log
$\frac{\text{Dice}}{2-\text{Dice}}$
Boundary Loss
Dice
Distance map weighted
Cross Entropy (CE)
Combo
One side
Two side distance map weighted
Hard mining
Down-weight easy examples
DiceFocal
Weight FP & FN
Multi-class
Tversky
GD
HD Loss
TopK loss
Focal loss
$\alpha+\beta=1$
Weight FP & FN
Asym.
pGD
DiceTopK
Focal Tversky
Distribution-based Loss
Compound Loss
Region-based Loss
Boundary-based Loss

后厂理工学院 houchangtech.com

Loss

- A collection of loss functions for medical image segmentation
- https://github.com/JunMa11/SegLoss

Cross-Entropy

Entropy

$$D_{KL}(p \parallel q) = H(p,q) - H(p)$$

Cross entropy

ce loss

$$Cross-entropy\ loss = -\sum_{c=1}^{M} Y \log(P)$$

bce loss

$$-(y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

Information Theory

- Probability and Information Theory
- http://www.deeplearningbook.org/contents/prob.html

Probability

- frequentist probability
- Bayesian probability

Bayes's Rule

- we knowP(y | x) and need to knowP(x | y)

$$P(\mathrm{x} \mid \mathrm{y}) = \frac{P(\mathrm{x})P(\mathrm{y} \mid \mathrm{x})}{P(\mathrm{y})}.$$

Conditional Probability

$$P(\mathrm{y} = y \mid \mathrm{x} = x) = \frac{P(\mathrm{y} = y, \mathrm{x} = x)}{P(\mathrm{x} = x)}.$$

What is Information

- Probability
- frequentist probability

self-information

- Likely events should have low information content, and in the extreme case,events that are guaranteed to happen should have no information contentwhatsoever.
- Less likely events should have higher information content.
- Independent events should have additive information. For example, findingout that a tossed coin has come up as heads twice should convey twice asmuch information as finding out that a tossed coin has come up as headsonce.

self-information

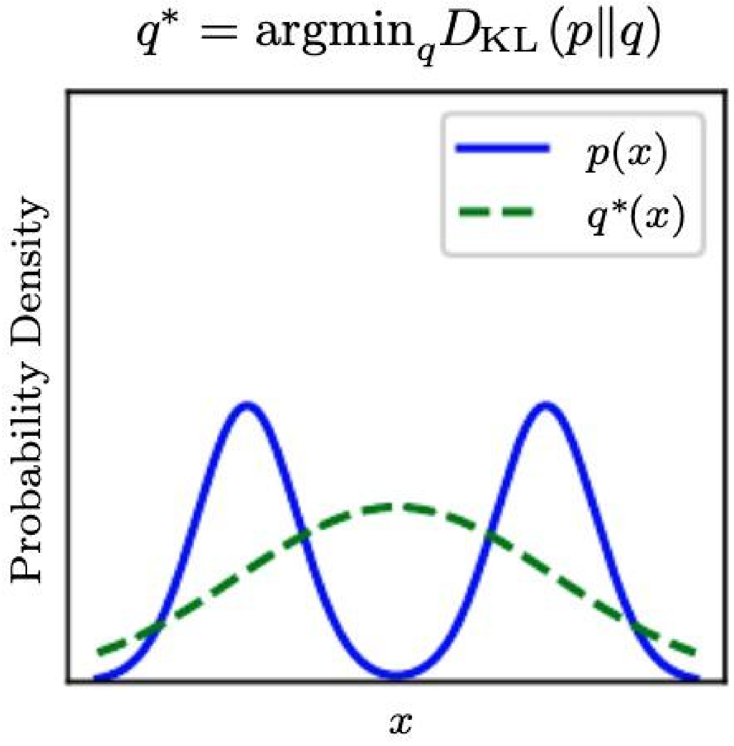
$$I(x) = -\log P(x).$$

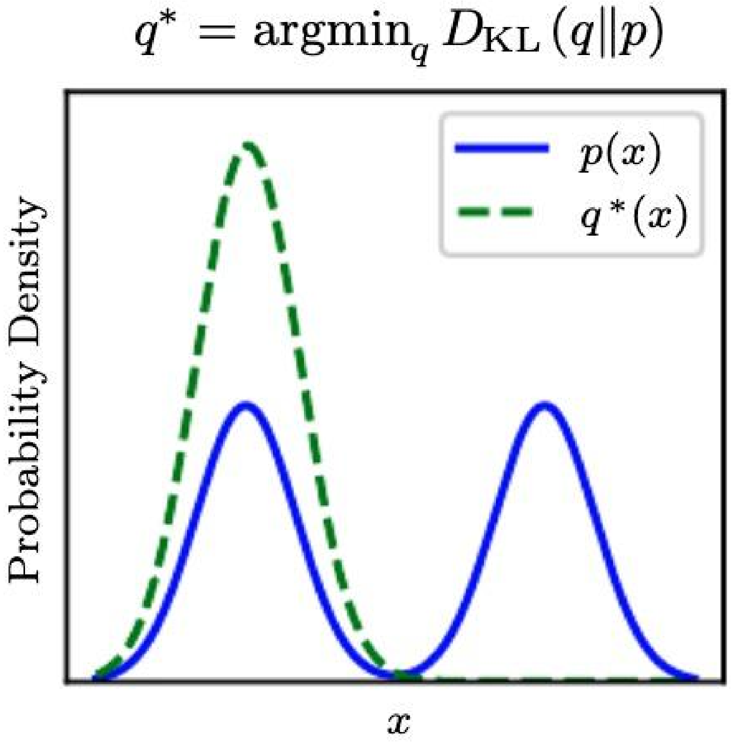
大道至简

Kullback-Leibler (KL) divergence

$$D_{\mathrm{KL}}(P\|Q) = \mathbb{E}_{\mathrm{x}\sim P}\left[\log \frac{P(x)}{Q(x)}\right] = \mathbb{E}_{\mathrm{x}\sim P}\left[\log P(x) - \log Q(x)\right].$$

Kullback-Leibler (KL) divergence





nll loss

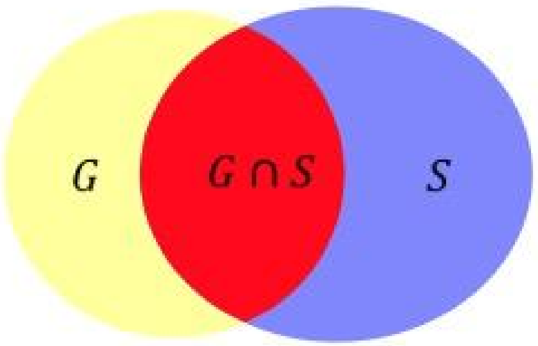
- cross-entroy loss

dice loss

- V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation
- https://arxiv.org/abs/1606.04797

Laplace smoothing

Dice



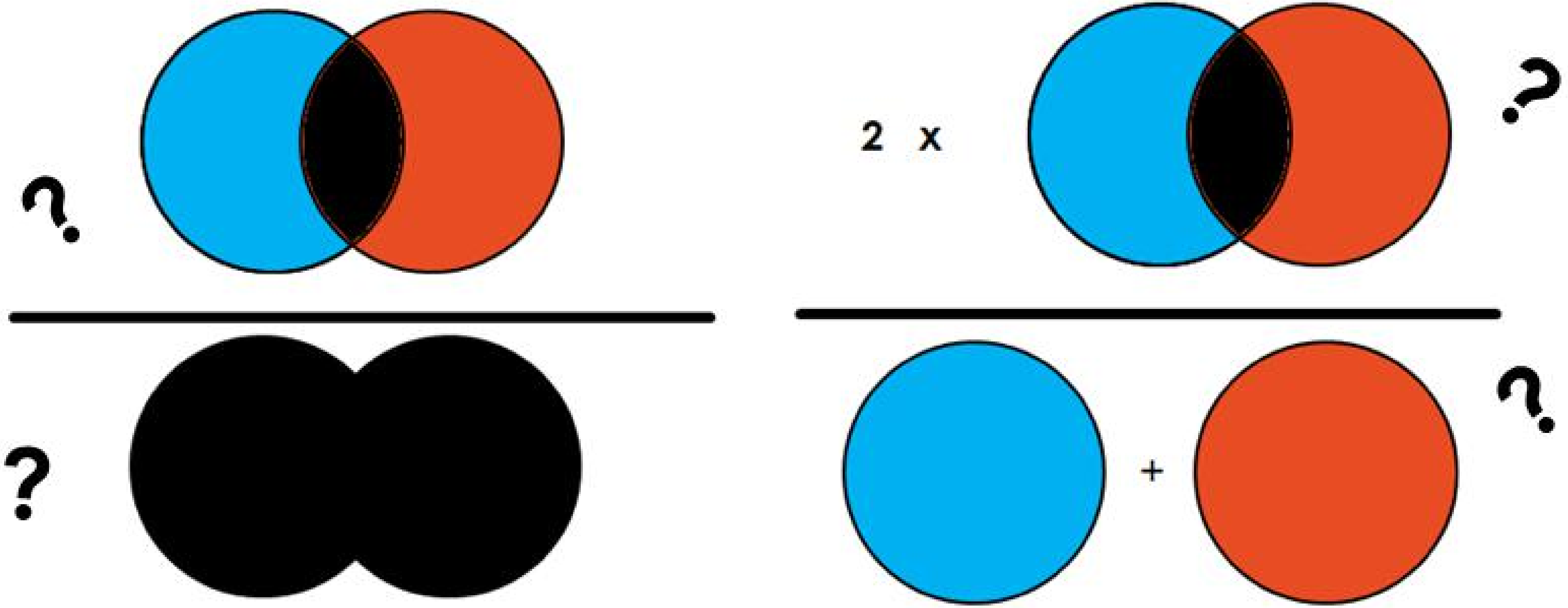
$$Dice\ loss = 1 - \frac{2|G \cap S|}{|G| + |S|}$$

Combinations

$$\mathrm{CE}(p, \hat{p}) + \mathrm{DL}(p, \hat{p})$$

Note that CE returns a tensor, while DL returns a scalar for each image in the batch. This way we combine local (CE) with global information (DL).

IoU and Dice Coefficient



Weight Initialization

- Initializing all weights to 0
- Initializing weights randomly

Weight Initialization

- Xavier
- Uniform
- Kaiming

Learning Rate

- Weight Initialization
- BN
- Adam...

Learning Rate

- Decay
- Cycle
- Warmup

后厂理工学院
houchangtech.com

Exponential Weighted Moving Average

$$v_0 = 0$$
$$v_1 = \beta v_0 + (1-\beta)\,\theta_1$$
$$v_2 = \beta v_1 + (1-\beta)\,\theta_2$$
$$v_3 = \beta v_2 + (1-\beta)\,\theta_3$$
$$\cdots$$

SGD+Momentum

- 手动挡
- 对学习率等超参数敏感

Adam

- 自动挡
- 最常用

课程总结

- mIoU的计算方法
- Loss的种类和组合
- Adam优化方法
- 调整学习率的算法和实现
- 单卡/多卡的训练方法

重难点

- mIoU
- dice loss
- Adam
- Cycle LR
- 单卡/多卡训练的实现

课程作业

- 使用Pytorch/Tensorflow实现项目训练

后厂理工学院

houchangtech.com

一所专注前沿互联网技术领域的创新实战大学