画像分類タスクにおける損失関数と注視領域の傾向調査

EP19013 伊藤天詞　　指導教授：山下隆義

1. はじめに

畳み込みニューラルネットワーク (CNN) による画像認識モデルの学習は，正解ラベルと予測確率の損失を最小化することである．CNN の学習には様々な損失関数が使用される一方で，それらの効果について深く議論されていない．そこで本研究では，各損失関数における認識精度，特徴表現，注視領域の 3 つの観点で傾向を体系的に調査する．

2. 損失関数

画像認識モデルの損失関数は,「出力空間に対するもの」と「特徴空間に対するもの」の 2 つに大別することができる．

2.1. 出力空間に対する損失関数

Cross Entropy Loss　Cross Entropy Loss (CE Loss) は，正解クラス確率を 1 に近づける損失関数である．CE Loss は，N をサンプルの総数，$\boldsymbol{y}$ を正解ラベルが 1 の確率分布，$\hat{\boldsymbol{y}}$ を予測確率分布として，式 (1) で表現される．

$$L_{CE}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i^T \cdot \log(\hat{\boldsymbol{y}}_i) \tag{1}$$

CE Loss は，$L_{CE}(\boldsymbol{y}, \hat{\boldsymbol{y}})$ が最小となるように学習する．

Complement Entropy Loss　Complement Entropy Loss (CoE Loss) は，Complement Objective Training [1] で提案された不正解クラス確率を平坦化する損失関数である．CoE Loss は，g を正解クラスのインデックスとして，式 (2) で表現される．

$$L_{CoE}(\hat{\boldsymbol{y}}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1, j\neq g}^{k} \frac{\hat{y}_{ij}}{1-\hat{y}_{ig}} \log\left(\frac{\hat{y}_{ij}}{1-\hat{y}_{ig}}\right) \tag{2}$$

CoE Loss は，$L_{CoE}(\hat{\boldsymbol{y}})$ が最大となるように学習する．CoE Loss を適用したモデルは CE Loss 最小化の後に CoE Loss 最大化を行うため，1 つのミニバッチで 2 回重みパラメータを更新する．

2.2. 特徴空間に対する損失関数

特徴空間に対する損失計算は，CE Loss と併せて用いる．

Center Loss　Center Loss は，特徴空間において，同じクラスの特徴量を近づける損失関数である．Center Loss は，入力画像と教師ラベルのペアを N サンプル含む集合 $D := \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$ に関して，学習可能なクラス重心を $\boldsymbol{w}^c$，特徴量を $\boldsymbol{f}_i$ として，式 (3) で表現される．

$$L_{Center} = \frac{1}{2}\sum_{i}^{N} \| \boldsymbol{f}_i - \boldsymbol{w}_{y_i}^c \|_2^2 \tag{3}$$

Center Loss は，対象クラスの特徴量とそのクラス重心との距離を最小化することで，クラス内分散を抑えることができる．

Prototype Conformity Loss　Prototype Conformity Loss (PC Loss) [2] は，特徴空間において，同じクラスの特徴量を近づけ，異なるクラスの特徴量を離す損失関数である．PC loss は，クラス数を k，正解クラスを y_i，学習可能なクラス重心 $\boldsymbol{w}^c$ と特徴量 $\boldsymbol{f}_i$ の関係を $\alpha_i = \|\boldsymbol{f}_i - \boldsymbol{w}_{y_i}^c\|_2$，$\beta_i = \|\boldsymbol{f}_i - \boldsymbol{w}_j^c\|_2$，$\gamma_i = \|\boldsymbol{w}_{y_i}^c - \boldsymbol{w}_j^c\|_2$ として，式 (4) で表現される．

$$L_{PC} = \sum_{i}^{N}\left\{\alpha_i - \frac{1}{k-1}\sum_{j\neq y_i}(\beta_i + \gamma_i)\right\} \tag{4}$$

PC Loss は，α_i を最小化することで特徴量をクラス重心に集め，$(\beta_i + \gamma_i)$ を最大化することでクラス間を離すように働く．

3. 評価実験

本実験では，CNN に CoE Loss，Center Loss，PC Loss を導入して傾向を調査する．

3.1. 実験概要

本実験では，認識精度だけでなく特徴空間を定量的評価，注視領域を定性的評価する．特徴空間の定量的評価には，クラス間の分離距離を評価する Silhouette score (Sil.) ，同じクラスの凝縮度合と異なるクラスの分散度合を評価する Calinski harabasz index (Cal.) を用いる．注視領域は，Grad-CAM を使用して可視化する．

3.2. 実験条件

ネットワークモデルは ResNet-20，活性化関数は ReLU，最適化手法はモーメンタム SGD，バッチサイズは 128，ドロップアウト率は 30%，学習の更新回数は 200 エポックである．また，事前学習として CE Loss を用いて 100 エポック学習する．学習用データセットは CIFAR-10 を用いる．

3.3. 実験結果

各損失関数における認識精度と特徴空間の定量的評価の比較を表 1 に示す．表 1 より，Center/PC Loss を導入すると認識精度，特徴空間のスコアが同等以上になることを確認した．また，CoE Loss + PC Loss の場合に最高精度となった．

表 1: 認識精度と特徴空間の定量的評価

		test acc.[%]	Sil.	Cal.
w/o CoE	CE	95.00	0.3773	2666.85
	Center	**95.32**	**0.7896**	**15588.98**
	PC	95.21	0.7877	15218.74
w/ CoE	CE	95.19	0.4653	3363.28
	Center	95.26	**0.7928**	14269.11
	PC	**95.36**	0.7924	**14280.56**

3.4. 注視領域の可視化

各損失関数におけるアテンションマップを図 1 に示す．図 1 より，Center/PC Loss を導入すると対象クラスの物体領域に注視することが確認できる．視覚的説明性の向上に併せて認識精度も同等以上になることから，Center/PC Loss を導入したモデルは画像内の小さな範囲に写る物体に対して有効であるといえる．

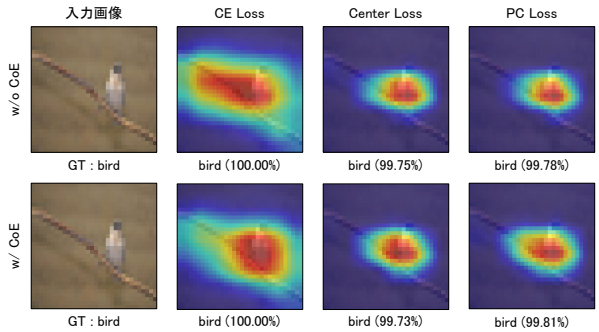

図 1: 各損失関数におけるアテンションマップの例

4. おわりに

本研究では，各損失関数における認識精度，特徴表現，注視領域の 3 つの観点で傾向を調査した．評価実験では，Center/PC Loss を導入した CNN は，物体を局所的に注視し，小さな範囲に写る物体に対して有効な特徴量を獲得できることが確認できた．今後は，Vision Transformer における傾向調査を行う予定である．

参考文献

[1] H. Chen, *et al.*, “COMPLEMENT OBJECTIVE TRAINING”, ICLR, 2019.

[2] A. Mustafa, *et al.*, “Adversarial Defense by Restricting the Hidden Space of Deep Neural Networks”, ICCV, 2019.