

# 自己教師あり学習のための適応的正負例データ拡張

宮井 淳行<sup>†</sup> 郁 青<sup>‡</sup> 伊神 大貴<sup>§</sup> 入江 豪<sup>¶</sup> 相澤 清晴<sup>||</sup>  
 東京大学<sup>†</sup> 東京大学<sup>‡</sup> NTT<sup>§</sup> NTT<sup>¶</sup> 東京大学<sup>||</sup>

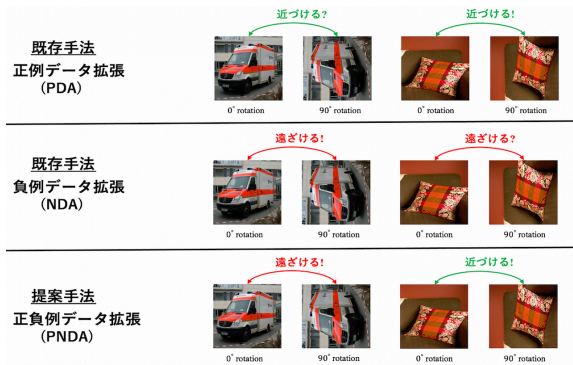


図1 既存手法 (NDA, PDA) と提案手法 (PNDA) の比較. 既存手法では、各画像の性質を考慮せずに、変形を正または負に扱うかを決定しているが、提案手法では各画像の性質に合わせて適切な方を選択している。

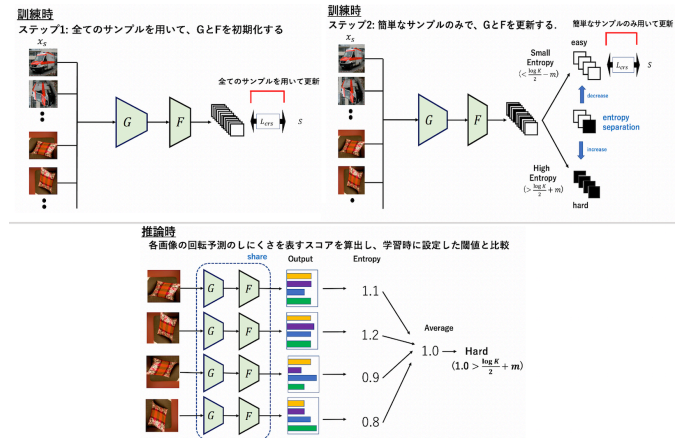


図2 RAISの訓練・推論の手順.

## 1 はじめに

自己教師あり学習におけるデータ拡張戦略には、正例データ拡張 (PDA) [1, 2] と負例データ拡張 (NDA) [3, 4, 5] がある。しかし、回転のような変形は PDA と NDA の両方に使われることがあり、そのような変形を正例・負例のどちらに指定するかは明確ではない [1, 5]。我々は、正例・負例のどちらに指定するかは、それぞれの画像の特徴に依存すると考えている。本論文では、回転に着目し、各画像の特徴に適応し、各画像に対して正負のどちらか良い方を適応的に選択し学習する適応的正負データ拡張 (PNDA) という新しいデータ拡張戦略を提案する。なお、正負例の適応的な選択を教師無しで学習する必要がある。これらのデータ拡張戦略を自己教師あり対照学習フレームワークに適用した結果、PDA と NDA が性能を低下させる可能性がある一方で、PNDA がベースラインを上回る性能を持つことを確認した。我々の研究は、各画像の特徴を考慮することの重要性を示している。

## 2 提案手法

図1に、提案手法であるPNDAの概要を示した。既存手法 [1, 2, 3, 4] は各画像の性質を考慮せずに変形を正または負のどちらか一方に指定しているが、提案手法のPNDAは

各画像の性質を考慮して正または負の最適な方を画像ごとに選択している。ただし、本研究でPNDAを行う際に回転によって特徴量分布が変化しない画像、rotation agnostic images (RAI) を全画像の中から選び出さなければならない。そこで、RAIを取り出すためのフレームワーク、RAIS (Rotation Agnostic Images Sampling) を提案する。そして、RAISにて各画像に対して回転を正負のどちらに用いるかを判断した後、PNDAを自己教師あり対照学習フレームワークに適用する方法を提案する。

### 2.1 RAIS (Rotation Agnostic Images Sampling)

90度の倍数、すなわち画像の0度、90度、180度、270度による回転を変形の集合  $S$  と定義する。全ての画像集合  $X$  に  $S$  を適用する。RAIは、回転させた画像の特徴分布が元画像の特徴分布と類似している。そのため、どのような変換  $y_s \in S$  が適用されるかを予測することが困難である。一方で、その他の画像の場合、回転した画像の特徴分布は元画像の特徴分布と異なる。そのため、どのような変換  $y_s \in S$  が適用されるかを予測することが容易となる。本セクションでは、本手法を分かりやすくするために、回転によって特徴量分布が変化する画像を  $X_e$  (easy to predict rotation)、回転によって分布が変化しない画像を  $X_h$  (hard to predict rotation) と呼ぶことにする。ここでの目的は、全てのサンプル  $X$  から  $X_h$  を抽出することである。

**3 ステップからなる訓練・推論の手順.** 図2に訓練・推論の手順を示した。Gは特徴量抽出器、Fは回転分類器を表している。我々は、全てのサンプル  $X$  から  $X_h$  を抽出するために、 $X_e$  に施された回転を正しく予測でき、 $X_h$  に施された回転は全く予測できないような回転分類器を作成する。これを達成するため以下に3stepからなる学習・推論の手順を示す。

Adaptive Positive or Negative Data Augmentation for Self-Supervised Learning

<sup>†</sup> Atsuyuki Miyai, The University of Tokyo

<sup>‡</sup> Qing Yu, The University of Tokyo

<sup>§</sup> Daiki Ikami, NTT

<sup>¶</sup> Go Irie, NTT

<sup>||</sup> Kiyoharu Aizawa, The University of Tokyo

**step1.** 最初に全サンプル  $X$  を用いて、回転分類器を学習する。その際に、分類器が訓練データに過学習する時点より前で step1 の学習を終えるようにする。

$$\mathcal{L}_{crs} = \frac{1}{B} \sum_{S \in \mathcal{S}} \sum_{x_S \in \mathcal{B}_S} -y_S \log(p(y^S|x_S)),$$

$$\mathcal{B}_S = \{S(x_i)\}_{i=1}^B \quad (1)$$

**step2.**  $X_e$  と  $X_h$  を分離するための回転画像分離損失を提案する。具体的には、まず以下の二つのロスを定義する。

$$\mathcal{L}_{es} = \frac{1}{B} \sum_{S \in \mathcal{S}} \sum_{x_S \in \mathcal{B}_S} \mathcal{L}_{es}(p(y^S|x_S)),$$

$$\mathcal{L}_{es}(p) = \begin{cases} -|H(p) - \rho| & (|H(p) - \rho| > m), \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

$$\tilde{\mathcal{L}}_{crs} = \frac{1}{B} \sum_{S \in \mathcal{S}} \sum_{x_S \in \mathcal{B}_S} \tilde{\mathcal{L}}_{crs}(p(y^S|x_S)),$$

$$\tilde{\mathcal{L}}_{crs}(p) = \begin{cases} \mathcal{L}_{crs}(p(y^S|x_S)) & (H(p) - \rho < -m), \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

式 (2) は、Entropy Separation Loss である。このロスによって、エントロピーが大きいもののエントロピーをより大きくし、小さいもののエントロピーをより小さくすることが可能となる。我々は、先行研究 [6] に従って、 $\rho = \frac{\log(K)}{2}$  と定義した。ここで、 $K$  はクラス数であり、ここでは、 $K = 4$  となる。また、 $m = 0.2$  と設定した。式 (3) にて、エントロピーが  $\rho - m$  より小さいものを  $X_e$  とみなして、 $X_e$  のみを用いて分類器を学習させる。ここで、エントロピーが  $\rho + m$  よりも大きいものは  $X_h$  とみなして、よりエントロピーを大きくすることで回転予測が全くできないようにする。最終的な、目的関数である回転画像分離損失はハイパーパラメータ  $\lambda$  を用いて以下となる。

$$\tilde{\mathcal{L}}_{crs} + \lambda \mathcal{L}_{es} \quad (4)$$

**step3.** 推論時には、ある画像に対して、4 種類の回転を施した画像を入力とし、それぞれの予測結果のエントロピーの平均値を各画像の回転予測のしにくさを表すスコアとする。そのスコアが学習時に  $X_h$  を判定するために設定した閾値  $\rho + m$  よりも大きいものを  $X_h$  として取り出す。

## 2.2 PNDA の自己教師あり対照学習への適用

PNDA では、RAI の場合、回転画像を正例データとして遠ざけるように学習し、RAI ではない画像の場合、回転画像を負例データとして遠ざけるように学習させる。PNDA の方法論をわかりやすくするために、本セクションでは、RAI を RP (Rotation Positive)、それ以外の画像を RN (Rotation Negative) と表す。

一般的な自己教師あり対照学習 [2] の場合、ある一枚の画像からクエリ画像とキー画像を作成し、クエリ画像とキー画像を近づけるように学習し、その他の画像を遠ざけるように学習させる。一方で、PNDA の場合、RP 画像については、クエリ画像と同じ画像から作られたキー画像に加え、キー画像を回転させた画像も正例画像として扱う。加えて、RN 画像については、キー画像のみを正例画像として扱い、キー画像を回転させた画像は負例画像として扱う。あるク

表 1 CIFAR100 を用いた線形識別の結果

手法	正例割合 (%)	精度 (%)
MoCo v2	-	67.51±0.08
MoCo v2 + PDA	100	63.36±0.12
MoCo v2 + NDA	0	67.28±0.32
MoCo v2 + PNDA (Ours)	12	<b>68.20±0.23</b>

表 2 Tiny ImageNet を用いた線形識別の結果

手法	正例割合 (%)	精度 (%)
MoCo v2	-	45.06±0.28
MoCo v2 + PDA	100	41.42±0.21
MoCo v2 + NDA	0	45.29±0.28
MoCo v2 + PNDA (Ours)	30	<b>46.35±0.1</b>

エリサンプル  $x_i$  に対して、 $x_i$  をネットワークに入れて得られる特徴量を  $z_i$ 、 $x_i$  の正例サンプルの集合を  $P_i$ 、負例サンプルの集合を  $N_i$  とした場合、温度パラメータ  $\tau$  を用いて目的関数は以下のように表せる。

$$\mathcal{L}_i^{\text{PNDA}} = -\frac{1}{|P_i|} \sum_{p \in P_i} \log \frac{\exp(z_i^\top z_p / \tau)}{\sum_{p \in P_i} \exp(z_i^\top z_p / \tau) + \sum_{n \in N_i} \exp(z_i^\top z_n / \tau)} \quad (5)$$

## 3 実験結果・結論

PNDA の評価を行うために、データセットには、CIFAR100 と ImageNet のサブセットである Tiny ImageNet の 2 つを使用した。ベースラインである対照学習のフレームワークには、MoCo v2 [2] を使用した。ネットワークのバックボーンには、CIFAR100 は ResNet50 [7]、Tiny ImageNet では ResNet18 [7] を使用した。表 1, 2 には、自己教師あり学習のための一般的な評価手法の一つである線形識別 (Linear evaluation) を行った結果を示した。これらの結果から、PDA と NDA は精度を下げてしまう可能性があるにも関わらず、PNDA はベースラインを超える性能を達成していることがわかる。この論文は、各画像の性質を考慮することの重要性について再考することを促し、より効果的な自己教師あり学習手法の開発につながることを期待する。

## 参考文献

- [1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.
- [2] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.
- [3] A. Sinha, K. Ayush, J. Song, B. Uzkent, H. Jin, and S. Ermon. Negative data augmentation. In *ICLR*, 2021.
- [4] S. Ge, S. Mishra, C.-L. Li, H. Wang, and D. Jacobs. Robust contrastive learning using negative samples with diminished semantics. *NeurIPS*, 2021.
- [5] J. Tack, S. Mo, J. Jeong, and J. Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *NeurIPS*, 2020.
- [6] Q. Yu, A. Hashimoto, and Y. Ushiku. Divergence optimization for noisy universal domain adaptation. In *CVPR*, 2021.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.