4Q-05

# 画像の周波数成分を考慮したデータ拡張による CNN の分布外データへの頑健性向上

向井皇喜　熊野創一郎　山崎俊彦

東京大学

## 1　はじめに

近年，畳み込みニューラルネットワーク（Convolutional Neural Network，以下 CNN）を用いた画像分類の精度は非常に高くなっており，人間と同等かそれ以上の精度を達成している．しかし一方で，人間には見えない意図的なノイズを加えた画像である adversarial examples によって CNN の認識が人間の認識と全く異なってしまったり，クラス分類とは関係のない，学習データとはかけ離れた分布外データに対して，高い確信度を持ってあるクラスに分類してしまうなどの問題が指摘されている [1, 2].

こうした人間の知覚とは異なる CNN の振る舞いに関し，入力画像の周波数に注目した研究がいくつかある．Wang ら [3] は，CNN は人間にとって知覚できる意味のある領域と，人間には知覚できない高周波の領域を利用して精度を高めているため，高周波にノイズがある adversarial examples などに対し，人間の知覚とは違った認識になることがあるとしていた．また，Guangyao ら [4] は，CNN は画像の周波数の振幅に敏感であるが，人間の知覚は画像の周波数の位相に敏感であることから，画像の位相に注目させるデータ拡張（Amplitude-Phase Recombination，以下 APR）を行った．それにより学習させたモデルは，adversarial examples や分布外データに対し，より頑健なモデルになることを示した．

本研究ではこれらを踏まえ，モデルに学習データの周波数領域の情報を学習させるため，次のような周波数領域でのデータ拡張の手法を提案し，それにより分布外データへの頑健性が上がることを示す．

Improving Robustness of CNNs to Out-of-Distribution Data by Data Augmentation Considering Frequency Components of Images
Koki Mukai, Soichiro Kumano, and Toshihiko Yamasaki
The University of Tokyo

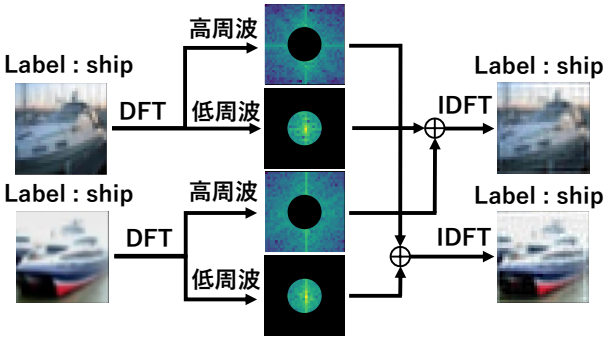


**図 1　RFC の概観図．同じクラスの画像を用いて周波数成分を高周波と低周波で取り換える．**

## 2　提案手法

画像を低周波成分と高周波成分に分解し，それを同じクラスの他の画像の高周波成分，低周波成分とそれぞれ組み合わせる（Replacement of Frequency Component，以下 RFC）ことにより新たな画像を生成する，というデータ拡張を行う．同じクラスごとに画像を組み合わせることによって，クラスごとの低周波，高周波の情報を保持しつつも，そのクラスの新たな画像を生成することができる．このデータ拡張の流れを図 1 に示し，以下に具体的な手順を示す．

CNN への入力画像を $\boldsymbol{x}$ とし，その周波数成分 $\boldsymbol{z}$ を，フーリエ変換 $\mathcal{F}(\cdot)$ を用いて $\boldsymbol{z} = \mathcal{F}(\boldsymbol{x})$ と表す．この $\boldsymbol{z}$ の低周波成分 $\boldsymbol{z}_l$，高周波成分 $\boldsymbol{z}_h$ を，周波数の中心（直流成分）のインデックスを $(c_i, c_j)$，半径 $r$ を用いて

$$\boldsymbol{z_l}(i,j) = \begin{cases} \boldsymbol{z}(i,j) & \left(\sqrt{(i-c_i)^2+(j-c_j^2)} < r\right) \\ 0 & \left(\sqrt{(i-c_i)^2+(j-c_j^2)} \geq r\right) \end{cases}$$
$$\boldsymbol{z_h}(i,j) = \begin{cases} 0 & \left(\sqrt{(i-c_i)^2+(j-c_j^2)} < r\right) \\ \boldsymbol{z}(i,j) & \left(\sqrt{(i-c_i)^2+(j-c_j^2)} \geq r\right) \end{cases} \quad (1)$$

とする．$\boldsymbol{x}$ と，$\boldsymbol{x}$ と同じクラスの画像 $\boldsymbol{x'}$ それぞれの低周波，高周波成分をそれぞれ $\boldsymbol{z_l}, \boldsymbol{z_h}, \boldsymbol{z'_l}, \boldsymbol{z'_h}$ とする

**表 1　各データ拡張によって CIFAR10 を学習したモデルの分布外データ（SVHN，LSUN，ImageNet，CIFAR100）への AUROC の比較．最も良い値を太字で示している．**

| method | Test acc.(%) | SVHN | LSUN | ImageNet | CIFAR100 |
|---|---|---|---|---|---|
| standard | 93.50 | 89.22 | 88.61 | 82.68 | 84.88 |
| cutmix [5] | 95.00 | 83.74 | 87.26 | 79.24 | 83.18 |
| mixup [6] | **95.31** | 82.91 | 87.41 | 76.63 | 78.09 |
| APR [4] | 95.21 | 98.13 | 92.94 | 84.46 | 88.45 |
| RFC (proposed) | 94.07 | 98.15 | 91.03 | 83.04 | 85.83 |
| RFC (proposed) & APR | 94.71 | **98.59** | **93.82** | **85.17** | **89.02** |

と，二つの画像間で高周波成分（低周波成分）を入れ替えた画像 $\boldsymbol{x}_{mix}, \boldsymbol{x}'_{mix}$ は，逆フーリエ変換 $\mathcal{F}^{-1}(\cdot)$ を用いて

$$\begin{aligned} \boldsymbol{x}_{mix} &= \mathcal{F}^{-1}(\boldsymbol{z_l}) + \mathcal{F}^{-1}(\boldsymbol{z'_h}) \\ \boldsymbol{x}'_{mix} &= \mathcal{F}^{-1}(\boldsymbol{z'_l}) + \mathcal{F}^{-1}(\boldsymbol{z_h}) \end{aligned} \tag{2}$$

と求められる．このようにして生成したデータ $\boldsymbol{x}_{mix}, \boldsymbol{x}'_{mix}$ を学習データに加える．

## 3　分布外データへの検証

実験の設定として，モデルは ResNet18，データセットは CIFAR10 を用い，最適化手法は SGD，学習率は 0.1 で始め，60 エポックごとに 0.2 倍に調整し，200 エポックまで学習させた．データ拡張の手法は，基本的な RandomCrop と RandomHorizontalFlip のデータ拡張を standard とし，それに加えて cutmix [5]，mixup [6]，APR [4] の手法を適用したデータ拡張，さらに APR [4] と RFC を両方適用したデータ拡張により，それぞれモデルを学習させた．

分布外データとしては SVHN，LSUN，ImageNet，CIFAR100 を用い，分布外データ検知にはモデルの最終層手前の softmax 層の出力の値を使い，分布外データかどうかの二値分類を行う．そしてその二値分類の評価指標として Area Under Receiver Operating Characteristic Curve（以下，AUROC）を用いる．各データ拡張の手法とそれによって学習したモデルの精度，AUROC の値を表 1 に示す．mixup [6] や cutmix [5] などの精度向上を目的とした空間領域のデータ拡張手法は，基本的なデータ拡張の standarad よりも精度は良いものの，分布外データへの頑健さを示す AUROC は下がっている．それに比べ周波数領域のデータ拡張である RFC は，精度の向上具合は他の手法に劣るが，分布外データに対してはより頑健になっていることが分かる．また，APR [4] と提案手法を組み合わせて使用することにより，さらに AUROC が上がることが確認できる．

## 4　まとめ

本研究では，画像の周波数領域を考慮したデータ拡張（RFC）を提案し，それにより学習したモデルは，空間領域でのデータ拡張の手法とは違い，分布外データへの頑健性が向上することが確認できた．今後はこの周波数領域のデータ拡張による adversarial examples などへの耐性の検証や，他のデータ拡張により学習したモデルとの，画像の注目領域の違いの検証などを考えている．

## 参考文献


[1] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. In *ICLR*, 2014.

[2] D. Hendrycks and K. Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2018.

[3] W. Haohan, W. Xindi, H. Zeyi, and X. Eric.P. High-frequency component helps explain the generalization of convolutional neural networks. In *CVPR*, pp. 8684–8694, 2020.

[4] G. Chen, P. Peng, L. Ma, J. Li, L. Du, and Y. Tian. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *ICCV*, pp. 458–467, October 2021.

[5] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *ICCV*, October 2019.

[6] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018.