Advanced Random Mix Augmentation : 深層学習による画像分類の性能向上のための 画像処理組み合わせの重複防止を用いたデータ拡張法*

林　鍾勳** ルイ笠原　純ユネス† 丸山　宏*** 淺間　一† 山下　淳†

Advanced Random Mix Augmentation :
Data augmentation method for improving performance of image classification
in deep learning using unduplicated image processing combinations

Jonghoon IM, Jun Younes LOUHI KASAHARA, Hiroshi MARUYAMA, Hajime ASAMA and Atsushi YAMASHITA

Data augmentation is a commonly used method for improving deep learning models in image classification. By adding slightly modified images that do not change the label of the original image to the training data set, the trained model becomes more robust against diverse characteristics of the input image. In this study, we propose a new data augmentation method by improving a previously-known random augmentation method. Our method consists of three steps; 1) determine the set of image modification operators and the number of augmented images, 2) determine the sequence of the image modification operators so that no duplicated sequences are generated, and 3) apply the sequence to augment images. The variety of augmentation is further increased by randomly determining the level (intensity) and the weight of combining the sequences. We applied our method on the CIFAR dataset and show that our method outperforms existing methods.

Key words: data augmentation, image processing, image classification, deep learning

1. 序　　論

深層学習を用いた画像分類における課題として過剰適合（Overfitting）がある．過剰適合は学習で利用された訓練データに対しては高性能であるが，学習してない未知のデータに対しては性能が落ちる現象である．一般的な深層学習は与えられたデータセットに対してリスク（エラー）を最小とする Empirical Risk Minimization（以下 ERM）に基づく学習 [1]であるため，過剰適合現象に対応することが難しい. もちろん, 学習の段階で特徴が異なる大量のデータを学習させることで過剰適合問題をある程度解決することができるが，常に大量のデータを確保しておくことは不可能である．上記の問題を解決する 1 つの方法として Vicinal Risk Minimization（以下 VRM）に基づく学習 [2]を利用する方法がある．VRM に基づく学習では，ERM のように与えられたデータセットのみを学習するのではなく，各入力データの近傍の分布まで学習を行う．一方，この VRM に基づく学習と非常に似たような方式がデータ拡張技術を用いて入力データを増やす方法である. そこで, 深層学習による画像分類における過剰適合をデータ拡張技術を用いて解決しようとする試みが増えている．

最近では画像処理を利用するデータ拡張技術が注目されている．この手法は原画像に対して画像処理をすることで決定境界（Decision boundary）付近で多量のデータを生成し，これらのデータを学習することで学習の性能を向上させる. また, この手法は処理が簡単であるため, 計算コストが低い長所があり, 近年多く利用されているデータ拡張手法である．

その代表的な手法として AugMix[3]がある．この手法は, 1 枚の画像に複数の画像処理を行い，生成された複数の画像を混ぜ合わせる手法である．しかし，AugMix ではまだ 2 つの課題が残っている. 1 つ目は, AugMix では画像処理後の変動の幅が小さく，多様な画像を生成するのに限界がある点である．2 つ目は，拡張画像を生成するため，画像処理の組み合わせをランダムに選択する過程で画像処理の組み合わせが重複される場合が発生し，拡張画像の多様性が低下する点である．上記の課題のうち, 前者の課題を解決するため，著者らは Random Mix Augmentation（以下 RMA[4]）を提案した．この手法は画像処理の強度をランダムに決めることで，より多様な拡張画像が生成できる．しかし，上記に述べた課題のうち，後者の課題は解決されていない．

そこで，本研究では既存の RMA を改善し，拡張画像の生成段階で画像処理の組み合わせが重複されることを防止することを目的とする. 拡張画像の多様性が低下することを予防し, より多様な拡張画像を生成することで，画像分類の性能を向上させる．この目的を達成するため，本研究では重複されないように画像処理の組み合わせを決める方法を提案する．

2. 関 連 研 究

従来のデータ拡張技術は Gaussian Blur, Brightness adjustment, ISO Noise, Filp, Rotation, Crop などを原画像に適用し，新たな学習データを生成する形式である．しかし，このような処理は処理前後の特徴がほぼ似ているので多様な特徴を持つ学習データを生成することは難しい．

機械学習に基づく手法として最適な Data augmentation policy を強化学習を用いて探索する AutoAugment[5]がある．この手法では 16 種類の Augmentation 手法を Search space として利用し良い

* 原稿受付 令和 4 年 5 月 26 日
掲載決定 令和 4 年 9 月 28 日
** 学生会員 東京大学大学院（東京都文京区本郷 7-3-1）
*** 東京大学大学院
† 正 会 員 東京大学大学院

性能を出すことができたが，非常に多くの計算コストおよび時間が必要である．その後，計算時間を短縮するため，Population Based Training アルゴリズム [6]を利用して Augmentation policy を探す Population Based Augmentation[7] や Tree-structured Parzen Estimator 方法 [8]で Augmentation policy を探す Fast AutoAugment[9] などの手法が提案された．一方，敵対的生成ネットワーク（Generative Adversarial Network：以下 GAN[10]）を用いたデータ拡張手法も提案されている．GAN は新しい画像を生成する生成者ネットワーク（Generator）とサンプルデータと生成者が生成した画像を区別する区別者ネットワーク（Discriminator）で構成される．生成者は区別者を騙しながら画像を上手く生成しようとし，区別者は与えられた画像が本物か偽物かを判別しながら互いに競争的に学習する．DCGAN[11]を利用して生成した Liver lesion image を学習に使用し，画像分類の性能を向上した方法 [12]や，CycleGAN[13]を用いて生成した顔の感情分類データを学習に使用し，Class imbalance を軽減して分類性能を向上した方法 [14]が提案された．また，1 枚の画像を GAN で学習して多くの拡張画像を生成する SinGAN[15]も提案されている．しかし，上記の機械学習に基づく手法や GAN に基づく手法は計算コストが高い問題がまだ残っている．特に GAN の場合はモード崩壊という Generator の学習が失敗し，同一の画像を生成してしまう問題も残っている．また，GAN を訓練するとき，訓練データ分布から外れた領域でのデータが得られない問題もある．

そこで，最近では処理が簡単で，計算コストも低い画像処理を利用する手法が注目されている．まず，画像内の一部領域をある値に変換する方式がある．その代表的な手法として，画像内に Random bounding box を利用して，その領域内の画素値を 0 にする Cutout[16]がある．また，同様な方式で，領域の画素値を Random noise, mean value, 0, 255 の中からいずれかに変換する Random Erasing Data Augmentation[17]も提案されている．

次に，2 つ以上の画像を混ぜ合わせる方式がある．2 つの画像とラベルを 0~1 の間の Lambda 値を利用して Weighted linear interpolation する Mixup[18]が代表的である．さらに，2 つの画像を混ぜ方を 8 種類に増やし，性能を改善させた手法 [19]や 4 枚の画像から Random crop した Patch を 1 枚に合成する手法 [20]も提案されている．

最後に，上記に述べた画像内の一部の領域をある値に変換する方式と 2 つ以上の画像を混ぜ合わせる方式を合わせた手法がある．Cutout と Mixup を融合した CutMix[21]がその代表的な例である．また，1 枚の画像に複数の Augmentation 技法を直列，並列に処理した後，それらを Random weight を用いて混ぜ合わせ，最後に原画像ともう一度混合する AugMix[4]がある．AugMix は上記の手法に比べて画像分類において高い性能を有する．しかし，序論で述べたように AugMix にはまだ 2 つの課題が残っている．1 つ目は，画像処理後の変動の幅が小さく，多様な画像を生成するのに限界がある点である．2 つ目は， 画像処理の組み合わせを用いて拡張画像を生成する過程で画像処理の組み合わせが重複され，拡張画像の多様性が低下する点である．

上記の問題を解決するため，著者らは RMA を提案した．RMA は大きく分けて 2 つの段階で構成されている．1 つ目は Augmentation 段階で，2 つ目は Mix 段階である．Augmentation 段階では，原画像に対して多様な画像処理を行い，複数の拡張画像を生成する．この段階で適用する画像処理の強度をランダムに決めることで，より多様な拡張画像が生成できるようにする．また，原画像に対して適用する画像処理の回数もランダムに決めることで，画像処理の組み合わせの数を増やし，画像合成後に生成できる拡張画像の範囲を広げる．Mix 段階では生成された複数の拡張画像をランダム重みを利用して合成することで，生成できる拡張画像の範囲を更に広げる．これより，決定境界付近でより多量の拡張画像を生成することで性能向上を達成した．しかし，既存の RMA では，Augmentation 段階で画像処理の組み合わせが重複され，拡張画像の多様性が低下する問題が解決されていない状況である．詳しくいうと 1 枚の合成画像を生成するためには複数の拡張画像が利用され，複数の拡張画像を生成するためには原画像に対して多様な画像処理の組み合わせが利用される．複数の拡張画像を生成する段階で全て同様な画像処理をして画像を合成することは，結局拡張画像 1 枚を生成することと同様な作業になり，このような現象が累積されると多様な画像を生成することが困難であり，処理時間およびシステムメモリの面でも非効率である．既存の RMA では各拡張画像の生成において画像処理の組み合わせの重複はほとんど考慮せず，単純に複数の拡張画像の生成に画像処理の組み合わせを利用することに重点を置いたため，手法の最適化という面ではまだ研究が進んでいなかった．

そこで，本研究では拡張画像の生成段階で画像処理の組み合わせが重複されることを防止する方法を提案することでこの課題を解決する．具体的には，予め発生可能な全ての画像処理の組み合わせを配列に保存して置き，その中から画像処理の組み合わせを非復元抽出することで，画像処理の組み合わせが重複されることを防止する．本研究では 1 回の画像生成において画像処理の組み合わせの重複を防ぐだけではなく，その以降の画像生成でも以前の画像生成で使用した画像処理の組み合わせと重複されないように画像処理の組み合わせを選択する新しいアルゴリズムを提案する．これによって多様な拡張画像の生成ができ，これらを学習することで既存の RMA に比べて画像分類において性能向上ができると期待される．

3. 提 案 手 法

3.1 コンセプト

前節で説明した既存の RMA の問題を解決するため，提案手法である Advanced Random Mix Augmentation（以下 ARMA）では，予め画像処理の組み合わせが重複されることを防止する処理を入れる．提案手法の処理フローを**図 1** に示す．

提案手法は大きく分けて 4 つの段階で構成される．1 つ目は Decision of combination size 段階，2 つ目は Combination selection 段階，3 つ目は Augmentation，4 つ目は Mix 段階である．Decision of combination size 段階では，生成する拡張画像の数および各拡張画像を生成するために適用する画像処理の回数を決める．Combination selection 段階では，以前の段階で決定された拡張画像の数および画像処理の回数の情報を基に実際に適用する画像処理の組み合わせを決める．各拡張画像の生成には画像処理の組み合わせが重複されないようにする．Augmentation 段階では，以前の段階で決まった画像処理の組み合わせの順番に画像処理を行い，複数の拡張画像を生成する．この段階で適用レベルというランダム要素を利用し，適用する画像処理のレベルおよび画像処理を行う回数をランダムに決めることで，画像処理の組み合わせの数を増やす．これによって，画像合成後に多様な拡張画像を生成できる．Mix 段階では生成された複数の拡張画像を重みを利用して合成する．このとき，合成に使用される重みは決まった範囲内でランダムに決める．これにより，多様な合成画像が

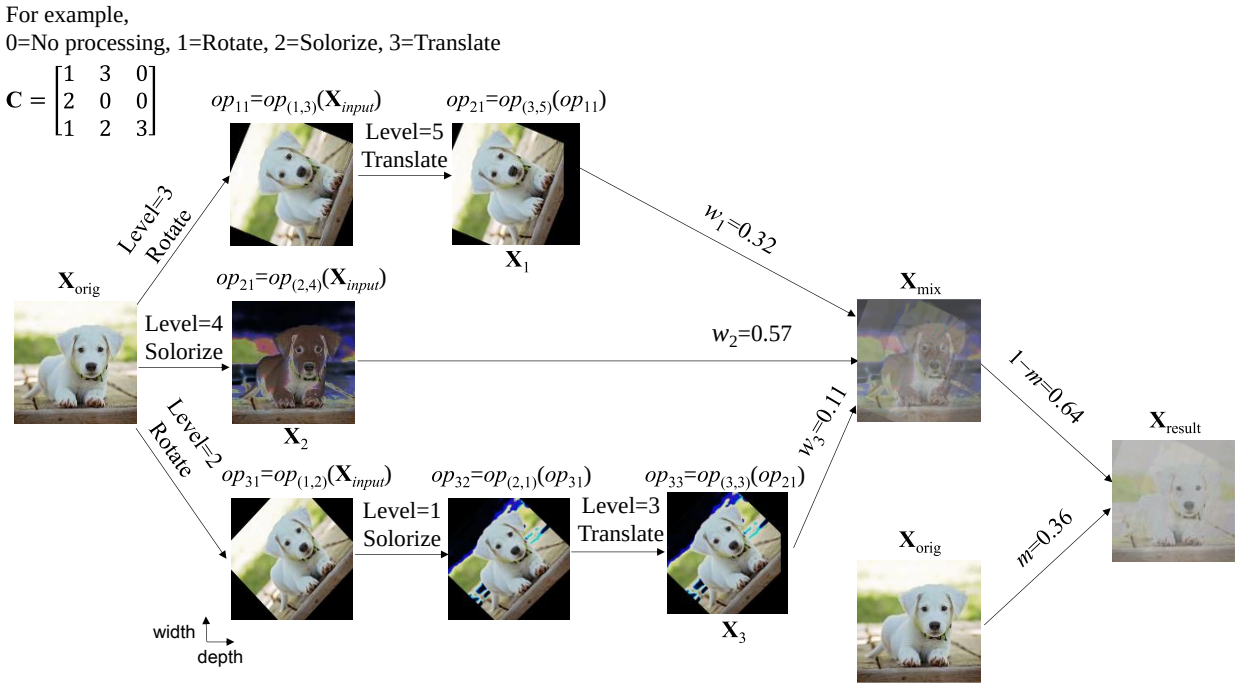

Fig. 1 Process flow of proposed method

生成できる．これは決定境界付近でより多量の拡張画像が生成できることを意味し，これらを学習することで学習の性能向上が期待される．

3.2 Decision of combination size

この段階では拡張画像の生成に使用する Combination Matrix の大きさを決める．Combination Matrix 式を式(1)に示す．

$$\mathbf{C}=\begin{bmatrix} a_{(1,1)} & \cdots & a_{(1,depth)} \\ \vdots & \ddots & \vdots \\ a_{(width,1)} & \cdots & a_{(width,depth)} \end{bmatrix}, \tag{1}$$

ここで，**C**は Combination Matrix を，width は生成する拡張画像の数を，depth は各拡張画像に適用する画像処理の回数を表す．Combination Matrix の大きさ depth×width となる．$a_{(i,j)}$は行列を構成する成分で，ここでは画像処理の種類を表す．図 1 のように width=3, depth=3 の場合は大きさ3×3の行列を生成する．これは，拡張画像を 3 枚生成し，各拡張画像では最大 3 種類の画像処理を適用する意味となる．

3.3 Combination selection

この段階では以前の段階で決めた Combination Matrix の各成分に画像処理技法を配置する．まず，画像処理の組み合わせの全てのケースを配列に保存して置く．例えば，図 1 のように depth=3 で，使用する画像処理の種類は 4 つ（Rotate=1，Solorize=2, Translate=3, Equalize=4）とし，処理なし=0 の場合も考慮すると，発生可能な全ての組み合わせ（重複可能）の数は重複組み合わせの公式により ${}_5H_3$= 35になる．提案手法では，より多様な各画像を生成するため，画像処理の組み合わせ内で同じ画像処理が 2 回以上重複実行されないようにする．発生可能な全ての組み合わせ（重複不可能）の数を表す式を式(2)に示す．

$$n_{all}=\sum_{i=1}^{depth} {}_m\mathrm{C}_i \ , \tag{2}$$

ここで，m は画像処理の数で n_{all} は組み合わせ（重複不可能）の数である．上記の例では発生可能な全ての組み合わせ（重複不可能）の数は処理をしない場合を除いて ${}_4C_1+{}_4C_2+{}_4C_3=14$ になる．1 つ目の項は 4 つの画像処理の中から 1 つを選ぶケースで，配列の形で表すと[1 0 0]，[2 0 0]，[3 0 0]，[4 0 0]となる．2 つ目の項は，4 つの画像処理の中から 2 つを選ぶケースで，配列の形で表すと[1 2 0]，[1 3 0]，[1 4 0]，[2 3 0]，[2 4 0]，[3 4 0]となる．3 つ目の項は 4 つの画像処理の中から 3 つを選ぶケースで，配列の形で表すと[1 2 3]，[1 2 4]，[1 3 4]，[2 3 4]となる．

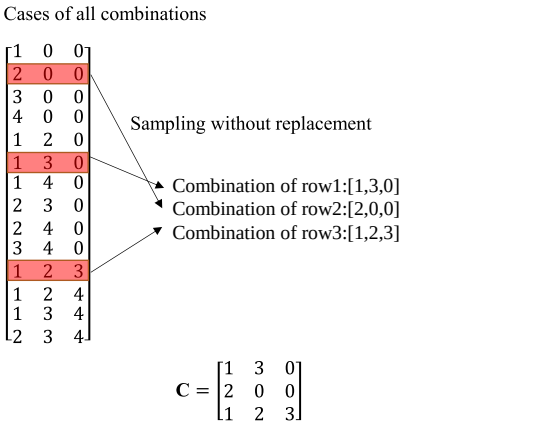

Fig. 2 Example of sampling without replacement

次に，生成する拡張画像の数ほど配列から画像処理の組み合わせを抽出する．ここでも，各拡張画像が同じ画像処理の組み合わせで生成されることを避けるため，配列から画像処理の組み合わせを非復元抽出する．**図 2** に配列から画像処理の組み合わせを非復元抽出する例を示す．図 2 の左側のように全ての画像処理の組み合わせを数値化して配列に順番に入れた後，その中から予め設定した Width ほど重複されないように抽出する．図 2 の例では Width=3 であるので配列から 3 つの画像処理の組み合わせ[1 3 0]，[2 0 0]，[1 2 3]を抽出した．その次に，抽出された画像処理の組み合わせを Combination Matrix の各行に入れる．その結果，図 2 の配列 **C** のように Combination Matrix が完成され，次の Augmentation 段階で実際の画像処理を行うとき，利用され

る．

この過程を Epoch ごとに繰り返す．配列から画像処理の組み合わせを全部抽出し，抽出できない場合は，配列をリセットして画像処理の組み合わせを抽出する．画像処理の組み合わせの数が Epoch 数に比べて少ない場合には，Epoch の間で画像処理の組み合わせが重複される場合が必ず発生する．このような場合でも多様な拡張画像を生成するため，提案手法では画像処理の処理強度であるレベルと画像合成で使用される重みをランダムにする．それぞれの詳細については 3.4 節と 3.5 節で説明する．

3.4 Augmentation

この段階では以前段階で決まった Combination Matrix の情報に従って原画像に対して画像処理（Operation）を適用し，複数の拡張された画像を生成する．入力画像に対する画像処理式を式(3)に示す．

$$\mathbf{X}_{output}=op_{(i,level)}(\mathbf{X}_{input}), \tag{3}$$

ここで，$\mathbf{X}_{input}$は入力画像，$\mathbf{X}_{output}$は出力画像，$op_{i,level}$は入力画像に適用する画像処理で以前段階で決まった Combination Matrix に従う．iは画像処理の種類を，$level$は画像処理の適用レベルを表す．レベルが大きくなるほど入力画像に対してより強い画像処理を適用する．例えば，図 1 の 1 番目の処理$op_{(1,3)}$では入力画像に対して Rotate をレベル 3 の強度で適用する．その次に，$op_{(3,5)}$処理では Translate をレベル 5 の強度で適用し，拡張画像 1 の生成を完了する．同様に別の拡張画像に対しても上記の処理過程を繰り返し，画像処理の組み合わせが重複されないように複数の拡張画像を生成する．以下に各画像処理の詳細について説明する．

3.5 Mix

この段階では Augmentation の結果により得られた各画像を合成する．拡張画像の多様性を増やすため，各画像の混合比はランダムに決める．式(4)に画像の合成式を示す．

$$\mathbf{X}_{mix}=\sum_{i=1}^{n} w_i\mathbf{X}_i,\ w_i\in\{0,1\} \tag{4}$$

ここで，$\mathbf{X}_i$は Augmentation 後の各画像，$\mathbf{X}_{mix}$は合成後の画像，w_iは画像合成時に使用される重みである．重みは$\sum_{i=1}^{n} w_i=1$となるように Dirichlet 分布上にあるn個の値をランダムに抽出する．各重みは 0 から 1 の間の実数値を持つ．

最後に得られた合成画像を原画像と合成することで，原画像の特徴をある程度維持させる．式(5)に画像の最終合成式を示す．

$$\mathbf{X}_{result}=m\mathbf{X}_{orig}+(1\text{-}m)\mathbf{X}_{mix},\ m\in\{0,1\} \tag{5}$$

ここで，$\mathbf{X}_{orig}$は原画像，$\mathbf{X}_{result}$は Mix の結果得られた画像，mは最終画像合成に利用される重みである．m は Beta 分布上にある 1 つの値をランダムに抽出する．m は 0 から 1 の間の実数値を持つ．

3.6 提案手法のまとめ

上記に述べたように提案手法提案手法は 4 つの段階で構成される．Decision of combination size 段階では拡張画像の生成に使用する Combination Matrix の大きさを決める．Combination selection 段階では画像処理の組み合わせの全てのケースを配列に保存して置く．次に各拡張画像が同じ画像処理の組み合わせで生成されることを避けるため，配列から画像処理の組み合わせを非復元抽出する．最後に抽出された画像処理の組み合わせを Combination Matrix の各行に入れる．Augmentation 段階では以前段階で決まった Combination Matrix の情報に従って原画像に対して画像処理を適用し，複数の拡張された画像を生成する．この段階で適用する画像処理のレベルおよび画像処理を行う回数をランダムに決めることで，画像処理の組み合わせの数を増やす．Mix 段階では生成された複数の拡張画像をランダム重みを利用して合成し，得られた合成画像を原画像ともう一度合成することで，原画像の特徴をある程度維持させる．

4. 実　　験

4.1 実験環境

実験では学習データとしては CIFAR データセットを利用した．これらは大きさ 32x32 のカラーデータで動植物や自然風景，乗り物などの画像が含まれている．CIFAR-10 は 10 クラスで，CIFAR-100 は 100 クラスで構成されている．データの数は CIFAR-10 の場合，各クラスあたり 5,000 枚の訓練データと 1,000 枚のテストデータで構成されており，CIFAR-100 の場合，各クラスあたり 500 枚の訓練データと 100 枚のテストデータで構成されている．CIFAR10-C および CIFAR100-C はそれぞれ CIFAR10 および CIFAR100 データに損傷（Corruption）を起こしたデータで Gaussian noise, Defocus blur, Snow, Jpeg compression, Brightness などの処理が利用される．

学習モデルとしては CNN[22], DenseNet[23], WideResNet[24]を利用した．Loss 関数は多様な入力範囲にかけて安定的である Jensen-Shannon Divergence Consistency Loss 関数 [25]を利用した．バッチサイズは 128，ラーニングレートは 0.1，Epoch は 100 と設定し，AugMix，RMA，ARMA では Depth=3，Width=3 でいずれも同じ値に設定し，RMA，ARMA で Level=5 と設定した．AugMix，RMA，ARMA の Augmetation 後，生成されたデータ数はいずれも訓練データ 50,000 枚×100epoch＝5,000,000 枚である．

Augmentation 段階では 9 種類（Auto contrast, Equalize, Posterize, Solarize, Rotate, Shear X, Shear Y, Translate X, translate Y）の画像処理を利用した．Autocontrast は画像内の特定の部分に集まっている値を全領域に均一に分布させる処理で，Equalize は画像のヒストグラムを均一化する処理である．Posterize は画像の階調（bit 数）を減らす処理で，Solarize は入力画像のある画素値が閾値より小さい場合はそのまま出力し，それ以外は画素値を反転する処理である．Rotate は画像を回転する処理，Shear は四角形の画像を平行四辺形に変換する処理，Translate は画像を平行移動する処理である．回転および移動処理によって生成された画像内の黒い部分は回転および移動以外の処理を行った他の拡張画像と合成され，合成重みによってその部分の情報が完全になくなるのではなく，その部分の画素値を減少させ，明るさを暗くすることになる．また，全ての拡張画像が回転および移動処理で生成されたとしても提案手法の最後の段階では，合成された拡張画像と入力画像をもう一度合成することで，画像内の黒い部分に該当する情報が完全になくなることを防止した．上記の 9 種類の画像処理は既存の AugMix で使用されている画像処理で RMA および ARMA でも同様な条件で AugMix と性能比較を行うため，上記の 9 種類の画像処理を利用した．**図 3** にレベル別に各 Augmentation を適用した結果の例を示す．

評価指標としては全データの中で誤分類された割合を示すエラーレートを利用した．エラーレートは式(6)で求められる．

$$E=\frac{(FP+TN)}{(TP+FP+TN+FN)}\times 100, \tag{6}$$

ここで，E はエラーレートを，TP は予測値と正解値が正であるもの，FP は予測値は正であるが正解値は負であるもの，TN は予測値は負であるが正解値は正であるもの，FN は予測値と正解値が負であるものの数を示す．

4.2　実験結果

実験では提案手法の性能を確認するため, データセットの数, 損傷有無，学習モデルによる各データ拡張法の性能比較を行った. 具体的には, 学習時に使用するデータの数による画像分類の性能評価をするため，CIFAR-10 と CIFAR-100 データセットを用いて性能を確認した. また, 学習時に使用していない未知のデータに対しては性能がどのくらいであるかを確認するため，CIFAR-10-C および CIFAR-100-C データセットを利用し，性能を確認した. 最後に提案手法が特定の学習モデルに依存せず, 良い性能を出しているのかを確認するため，CNN，DenseNet，WideResNet を用いて性能を確認した.

4.2.1　データセットの数による性能比較

CIFAR-10，CIFAR-10-C，CIFAR-100 そして CIFAR-100-C データセットに対して Standard（データ拡張を使用せずに Wide Resnet のみで学習），Cutout，Mixup，CutMix，AugMix，RMA そして提案手法である ARMA で学習を行い，画像分類におけるエラーレートの比較実験を行った．その結果を**図 4** に示す．グラフの横軸は使用した CIFAR データセットの種類を，縦軸はエラーレートを, 棒グラフの各色は使用したデータ拡張法を表す. 具体的なエラーレートの数値は**表 1** に示す．実験結果，CIFAR-10 の場合は Mixup が 4.15%で一番エラーレートが低かったが，それ以外の CIFAR-10-C，CIFAR-100 そして CIFAR-100-C データ

Fig. 3　Example of applying augmentation to each level
(In order to make it easier to see the changes by level, we emphasized each image processing rather than the original image processing)

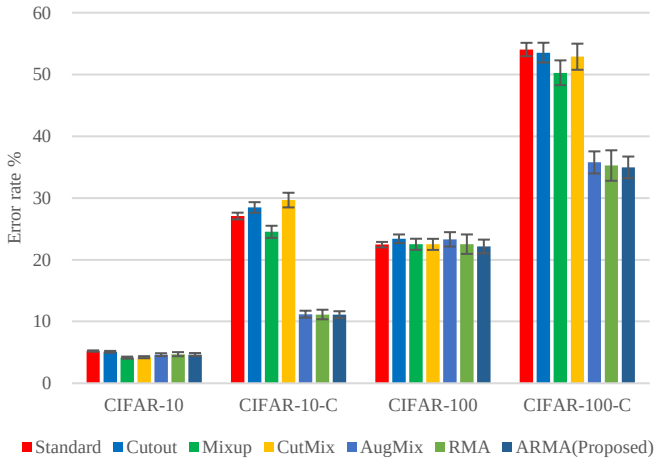

Fig. 4 Error rates of various methods on CIFAR dataset using a WideResNet

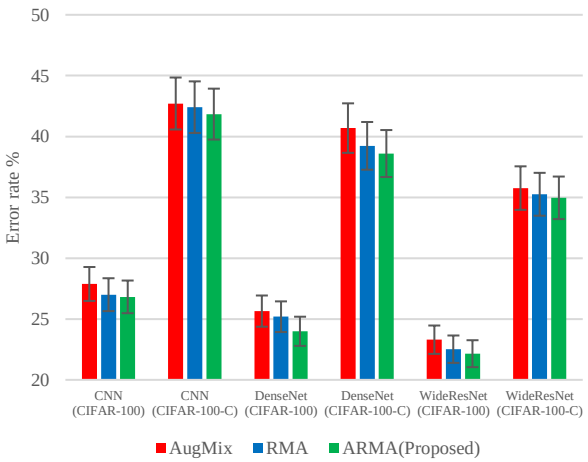

Fig. 5 Error rates of various methods on CIFAR dataset using CNN, DenseNet and Wide ResNet

Table 1 Error rates of various methods on CIFAR dataset using a WideResNet

Dataset	Standard	Cutout[16]	Mixup[18]	CutMix[21]	AugMix[3]	RMA[4]	ARMA(Proposed)
CIFAR-10	5.22	5.09	**4.15**	4.22	4.63	4.71	4.66
CIFAR-10-C	27.08	28.47	24.52	29.66	11.19	11.13	**11.11**
CIFAR-100	22.44	23.39	22.50	22.49	23.30	22.52	**22.15**
CIFAR-100-C	54.04	53.53	50.27	52.87	35.76	35.25	**34.96**

Unit %

Table 2 Error rates of various methods on CIFAR dataset using CNN, DenseNet and Wide ResNet

Model	Dataset	AugMix	RMA	ARMA (Proposed)
CNN[22]	CIFAR-100	27.88	27.00	**26.82**
	CIFAR-100-C	42.71	42.41	**41.84**
DenseNet[23]	CIFAR-100	25.65	25.19	**23.99**
	CIFAR-100-C	40.69	39.23	**38.60**
WideResNet[24]	CIFAR-100	23.30	22.52	**22.15**
	CIFAR-100-C	35.76	35.25	**34.96**

Unit %

セットの場合は提案手法がそれぞれ 11.11%，22.15%、34.96%で他手法に比べて一番エラーレートが低かった．特に，CIFAR-10 および CIFAR-10-C より CIFAR-100 および CIFAR-100-C データセットを利用した場合が，他の手法に比べて性能向上の効果が大きかった．CIFAR-100 データセットを基準として Standard，Cutout，Mixup，CutMix，AugMix，ARMA，RMA の順番に分散が大きくなった．これは，Standard から Cutout，Mixup，CutMix，AugMix，ARMA，RMA の順番にランダム性が大きくなるためであると考えられる． RMA と ARMA ではバリエーションの差が 2%程度で，この値ほど精度の差があった．RMA は画像処理の組み合わせの重複が可能であり，発生可能な画像処理の組み合わせの数が ARMA より多いため，ランダム性が ARMA より大きくなり，分散も ARMA より大きくなったと考えられる．また，発生可能な画像処理の組み合わせの数は ARMA より RMA の方が大きかったが，重複された画像処理の組み合わせが含まれていたため，実際の精度は ARMA が RMA より高かったと考えられる．

実験結果に対する考察として AugMix，RMA，提案手法は 1 枚の画像に多様な画像処理を適用して拡張画像を生成する手法であるため，データの数が少ない場合，画像分類の性能向上の効果が大きい．そのため，各クラスあたりデータの数が 500 個である CIFAR-100 データセットの場合は AugMix，RMA，提案手法が他手法に比べて非常に高い性能であったと考えられる．特に 提案手法の画像分類における精度が一番高かったが，これは提案手法には AugMix と RMA には存在しない画像処理の重複を防止する処理が入っているためであると考えられる．

一方，各クラスあたりデータの数が 5,000 個である CIFAR-10 データセットの場合は Mixup の性能が一番高かった．この結果に対する考察としてデータセットの数が十分に確保されている場合には様々な画像処理を用いて多様な拡張画像を生成しても既にそのようなデータが確保されている可能性があるため，大きい性能向上はできないと考えらあれる．データの数が十分に確保されている場合には，既存のデータセットを単純に合成して各クラス間の識別力を高める Mixup のような手法が画像分類においてより効果的であると考えられる．

表 1 の性能比較の結果を見ると CIFAR-100 データセットの場合， AugMix のエラーレートは 23.30%，RMA のエラーレートは 22.52%，ARMA のエラーレートは 22.15%であった．AugMix より RMA のエラーレートが 1.15%低いことからランダム性が学習の性能を向上させたことが確認できた．また，RMA より ARMA のエラーレートが 0.37%低いことから画像処理の組み合わせの重複防止が学習の性能を向上させたことが確認できた．まとめると，画像分類においてランダム性は 1.15%，重複防止は 0.37%の性能向上に寄与した．

4. 2. 2　データセットの損傷有無による性能比較

CIFAR-10 および CIFAR-100 より CIFAR-10-C および CIFAR-

Table 3 Error rates at each image processing level on CIFAR dataset using a WideResNet

Dataset	Level=1	Level=2	Level=3	Level=4	Level=5	Level=6	Level=7
CIFAR-100	22.77	22.51	22.46	22.43	**22.15**	22.50	23.06
CIFAR-100-C	39.56	37.01	36.14	35.44	**35.34**	35.39	35.44

Unit %

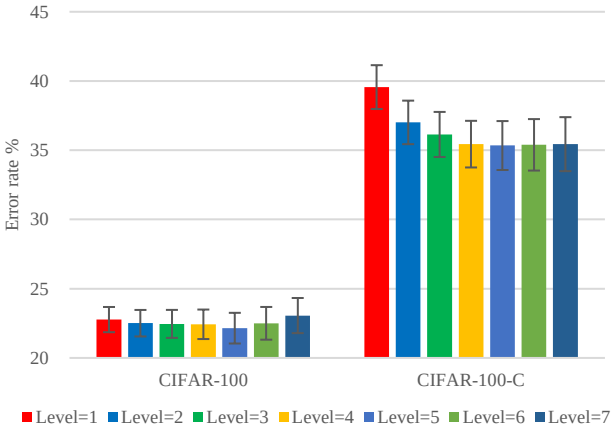

Fig. 6 Error rates at each image processing level on CIFAR dataset using a WideResNet

Table 4 Number of duplicate image processing in RMA and ARMA

Number of runs	1,000,000	2,000,000	3,000,000	4,000,000	5,000,000
RMA	22,895	46,311	69,453	92,860	116,412
ARMA	0	0	0	0	0

100-C データセットを利用した場合, エラーレートの低下が大きかった. 特に, AugMix, RMA, 提案手法のエラーレートが低かったが, これは上記の手法が他手法より多様な画像処理を使用して拡張画像を生成することで入力データだけではなく未知のデータに対して画像分類の性能を高めるためであると考えられる. また, RMA および提案手法では画像処理の強度をランダムに決める処理が追加され, AugMix より多様な拡張画像の生成ができ, より高性能であると考えられる.

一方, AugMix および RMA では拡張画像を生成する場合, 画像処理の組み合わせが重複される場合が発生する. これに対して提案手法の場合は, 画像処理の組み合わせの数が Epoch 数より多い場合は画像処理の組み合わせの重複が発生しないし, 画像処理の組み合わせの数が Epoch 数より少ない場合は, 画像処理の組み合わせの重複を最小限に発生するように処理を行う. そのため, 提案手法は AugMix および RMA より多様な拡張画像の生成ができ, 画像分類の性能を向上させることができたと考えられる.

4.2.3 学習モデルによる性能比較

次に, 上記の実験で他手法に比べて性能が格段に良い 3 つのデータ拡張手法（AugMix, RMA, 提案手法）に対して学習モデルを変えながら性能の比較を行った. その結果を**図 5** に示す. グラフの横軸は使用した学習モデルおよび CIFAR データセットの種類を, 縦軸はエラーレートを, 棒グラフの各色は使用したデータ拡張法を表す. 具体的なエラーレートの数値は**表 2** に示す.

実験結果, CNN, DenseNet, WideResNet いずれも提案手法が他手法より画像分類の性能向上が大きかった. これは, 提案したデータ拡張方法が学習と独立に行われるためであると考えられる. 学習モデルに依存せずにデータ拡張を行うため, 前節で述べたように画像処理におけるランダム強度および画像処理の組み合わせの重複防止を利用した提案手法が AugMix および RMA より画像分類において高性能であったと考えられる.

4.2.4 Augmetation の画像処理の最大強度による性能比較

Augmetation の画像処理の最大強度による学習性能比較の結果を**図 6** および**表 3** に示す. 実験結果, 画像処理の最大強度が多くなるほどエラーレートは低くなり, Level=5 で一番低かったが, その以降は最大強度が多くなるほどエラーレートは高くなった. この原因としては Equalize および Posterize 処理の場合, 画像処理の強度が 6, 7 になると本文中の図 3 のように画像の特徴がなくなり, 適切に学習ができなかったと考えられる.

4.2.5 画像処理の組み合わせの重複回数比較

画像処理の組み合わせの重複防止処理が含まれていない RMA と重複防止処理が含まれている ARMA の画像処理の組み合わせの重複回数を比較した結果を**表 4** に示す. この実験で比較した重複は図 2 における Combination Matrix の各行の重複の数である. 今回の実験では総 500 万枚の画像を生成し, 生成の途中で画像処理の組み合わせが重複された回数を数えた. 画像を 500 万回生成する間に重複が発生した回数は RMA の場合は 116,412 回で ARMA の場合は 0 回であった. これは表 1 の性能比較の結果と合わせると, 提案手法による画像処理の組み合わせの防止が画像分類において性能向上に有効であることを意味すると考えられる.

5. 結　　論

本研究では, 既存の Random Mix Augmentation を改善し, 拡張画像の生成段階で画像処理の組み合わせが重複されないようにすることで, 拡張画像の多様性が低下することを防止する手法を提案した. 実験結果により, 提案手法は既存の手法より学習データの数が少ないほど性能向上の幅が大きかった. また, 学習時に使用してない未知のデータセットに対しても他手法に比べて画像分類において高性能であった. 最後に, 提案手法は特定の学習モデルに依存せず, 実験で使用した全ての学習モデルにおいて他手法より高性能であった. 上記の実験結果より, 提案手法の重複防止処理を用いたデータ拡張手法の効果が示された. しかし, データの数が十分に確保されている場合は大きい性能向上はできなかった.

今後の課題としてデータの数が十分に確保されている場合にも大幅に性能向上ができるように手法を改善する必要がある. また, 学習データに対する適切な画像処理の選び方および適用する画像処理の強度の決め方についての検討が必要である.

参　考　文　献

1) V. N. Vapnik and A. ya. Chervonenkis: On the Uniform Convergence of Relative Frequencies of Events to their Probabilisties, Theory of Probability and its Applications, (1971) 264.

2) O. Chapelle, J. Weston, L. Bottou and V. Vapnik: Vicinal Risk Minimization, In Proceedings of the 13th International Conference on Neural Information Processing Systems, (2000) 395.
3) D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer and B. Lakshminarayanan: Augmix: A Simple Data Processing Method to Improve Robustness and Uncertainty, In Proceedings of the International Conference on Learning Representations, (2020).
4) 林 鍾勲, ルイ笠原 純ユネス, 山下 淳, 淺間 一: Random Mix Augmentation: 深層学習による画像分類の性能向上のためのデータ拡張法, 動的画像処理実利用化ワークショップ 2022 講演論文集 (DIA2022), (2022) 414.
5) E. D. Cubuk, B.Zoph, D. Mane and V. Vasudevan: AutoAugment: Learning Augmentation Policies from Data, arXiv preprint arXiv:1805.09501, (2018).
6) M. Jaderberg, V. Dalibard, S. Osindero, W. M. Czarnecki, J. Donahue, A. Razavi, O. Vinyals, T. Green, I. Dunning, K. Simonyan, C. Fernando and K. Kavukcuoglu: Population Based Training of Neural Networks, arXiv preprint arXiv:1711.09846, (2017).
7) D. Ho, E. Liang, X. Chen, I. Stoica and P. Abbeel: Population Based Aumentation: Efficient Learning of augmentation Policy Schedules, In International Conference on Machine Learing, (2019) 2731.
8) J. Bergstra, R. Bardenet, Y. Bengio and B. Kegl: Algorithms for Hyper-parameter Optimization, Advances in Neural Information Processing Systems, (2011) 2546.
9) S. Lim, I. Kim, T. Kim, C. Kim and S. Kim: Fast AutoAugment, arXiv preprint arXiv:1905.00397, (2019).
10) I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair and Y. Bengio: Generative Adversarial Nets, Advances in Neural Information Processing Systems 27, (2014) 2672.
11) A. Radford, L. Metz and S. Chintala: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv preprint arXiv:1511.06434, (2015).
12) J. Y. Zhu, T. Park, P. Isola and A. A. Efros: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, In Proceedings of the IEEE International Conference on Computer Vision, (2017) 2223.
13) M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger and H. Greenspan: GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, arXiv preprint arXiv:1803.01229, (2018).
14) X. Zhu, Y. Liu, Z. Qin and J. Li: Data Augmentation in Emotion Classification Using Generative Adversarial Networks, arXiv preprint arXiv:1711.00648, (2017).
15) T. R. Shaham, T. Dekel and T. Michaeli: SinGAN: Learning a Generative Model from a Single Natural Image, In Proceedings of the IEEE International Conference on Computer Vision, (2019) 4570.
16) T. Devries and G. W. Taylor: Improved Regularization of Convolutional Neural Networks with Cutout, arXiv preprint arXiv:1708.04552, (2017).
17) Z. Zhong, L. Zheng, G. Kang, S. Li and Y. Yang: Random Erasing Data Augmentation, In Proceedings of the AAAI Conference on Artificial Intelligence, **34**, 7 (2017) 13001.
18) H. Zhang, M. Cisse, Y. N. Dauphin and D. Lopez-Paz: Mixup: Beyond Empirical Risk Minimization, arXiv preprint arXiv:1710.09412, (2017).
19) C. Summers, and M. J. Dinneen: Improved Mixed-Example Data Augmentation, In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision(WACV), (2019) 1262.
20) R. Takahashi, M. Takashi and U. Kuniaki: Data Augmentation using Random Image Cropping and Patching for Deep CNNs, IEEE Transactions on Circuits and Systems for Video Technology, **30**, 9 (2019) 2917.
21) 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 Proceedings of the IEEE International Conference on Computer Vision, (2019) 6023.
22) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel: Backpropagation applied to handwritten zip code recognition, Neural computation, **1**, 4 (1989) 541.
23) G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Winberger: Densely Connected Convolutional Networks, arXiv preprint arXiv:1608.06993, (2018).
24) S. Zagoruyko, and N. Komodakis: Wide Residual Networks, In Proceedings of the British Machine Vision Conference, (2020) 87.
25) Q. Xie, Z. Dai, E. Hovy, M. Luong and Q. V. Le: Unsupervised Data Augmentation for Consistency Training, In Proceedings of the International Conference on Learning Representations, arXiv preprint arXiv:1904.12848, (2019).