

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

○ 林 鍾勳†, ルイ笠原 純ユネス†, 山下 淳†, 浅間 一†

○ Jonghoon IM †, Jun Younes LOUHI KASAHARA†, Atsushi YAMASHITA †

and Hajime ASAMA †

†: 東京大学, im@robot.t.u-tokyo.ac.jp

<要約> 画像分類において深層学習モデルの性能を向上させるため、一般的に利用される手法の 1 つはデータ拡張である。より多様な特徴を持つデータセットを原本データセットに加えることで、ある程度の変化があるデータに対しても対応できるようになる。本研究では、深層学習モデルの性能を高めるため、データを拡張する手法を提案する。まず、原画像に対して多様な画像処理を施し、複数の画像を生成する。この段階で適用レベルというランダム要素を導入する。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。これにより、決定境界付近のデータを多量に生成することができ、これらのデータを学習することで学習の性能向上が期待される。その次に、生成された複数の画像をランダム重み利用し合成することで、最終的に 1 枚の画像を生成する。実験結果により、提案手法は既存の手法より画像分類において性能を向上させることが示された。

<キーワード> Data augmentation, Image processing, Image classification, Deep learning

1. 序論

現在、深層学習技術はコンピュータビジョン、自律走行、翻訳、音声認識などの多様な分野で利用されている。特にコンピュータビジョン分野では、深層学習技術を活用し、オブジェクトを分類、認識、分割そして追跡する研究が活発に行われている。

一般的に、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）基板の学習である[1]。これにより、深層学習モデルは学習したデータセットに対しては高性能であるが、新に入力されたデータセットが学習時に利用されたデータセットと属性が違う場合、一般的には性能が落ちる。このような現象を過剰適合（Overfitting）と呼ぶ。過剰適合を減らす最良の手法は学習過程で大量のデータを学習することである[2]。ここで学習用データは多様な特徴を含むべきである。しかし、実際には多様な特徴を持つ多量のデータを確保することが難しい場合が多い。上記の問題を解決するため、提案された手法が Vicinal Risk Minimization（以下 VRM）基板の学習で

ある[3]。この方式では与えられたデータセットのみではなく、そのデータの近傍の分布まで活用して学習を行う。あるデータの近傍の分布を取得するとき、よく使われる手法がデータ拡張である。すなわち、多様な特徴を持つ多量のデータを収集することが難しい場合は深層学習の性能向上のためにデータ拡張技術の使用が必要である。

データ拡張手法は大きく分けて 3 つの種類で分類される。1 つ目は古典的な画像処理を利用する手法である。この手法では、回転、平行移動、ぼかしなどの処理を画像に適用し、学習データとして利用する。しかし、このような処理は処理前と処理後の特徴がほぼ似ているので多様な特徴を学習させることが難しい。

2 つ目は敵対的生成ネットワーク（Generative Adversarial Network : 以下 GAN）を利用する手法である[4]。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

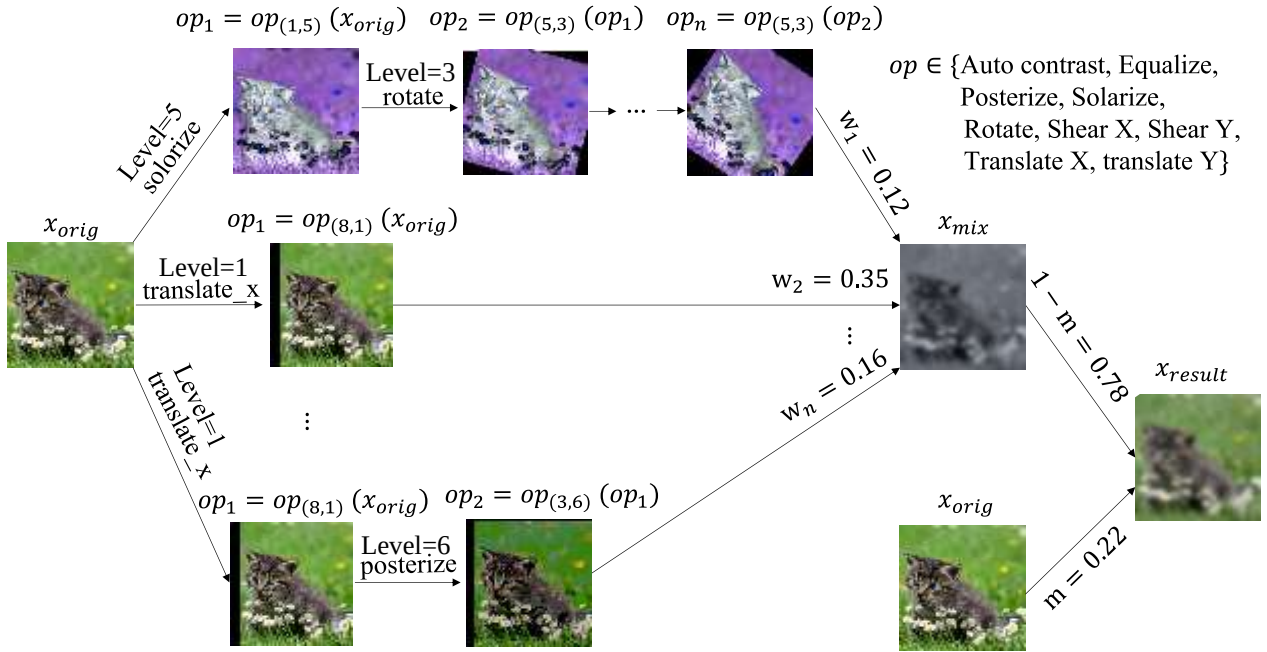


図1 提案手法の処理フロー

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。しかし、この手法は計算コストが高く、また、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が残っている。

3 つ目は画像処理と画像合成を利用する手法である。この手法は原画像に対して画像処理および画像合成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。これは画像を合成することが VRM 基板の学習をすることと同様な効果を出すためである[6]。

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。画像の一部を切り取る Cutout[5]、2枚の画像を合成する Mixup[6]、Cutout と Mixup を融合した CutMix[7]などの手法が提案されているが、複数の変換をかけた画像を最後に混ぜ合わせる手法である AugMix[8]が一番良い性能を出していると知られている。しかし、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。

本研究では、既存の画像処理と画像合成を利用した AugMix に適用レベルというランダム要素を加えて決定境界付近でより多量の拡張画像を生成し、これらを学習することで既存の手法より高性能なデータ拡張手法を提案することを目的とする。

2. 提案手法

2.1 提案手法のコンセプト

既存の AugMix ではそれぞれの画像に対する画像処理の種類およびそれぞれの画像の混合比のみをランダムに決定する。これに対し、提案手法では適用する各画像処理の種類に適用レベルをパラメータとして追加し、そのレベルもランダムに決めることで、より多様な画像が生成できるようにする。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2つのステップに構成される。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導入する。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 (Augmented image) 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も定めた範囲内でランダムに決める。これにより、同じ画像処理を行ったとしても重みを異なる値に設定し、合成することで、異なる合成画像が生成できる。合成に使用する重みをランダムに決めることで、より多様な合成画像が生成できる。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

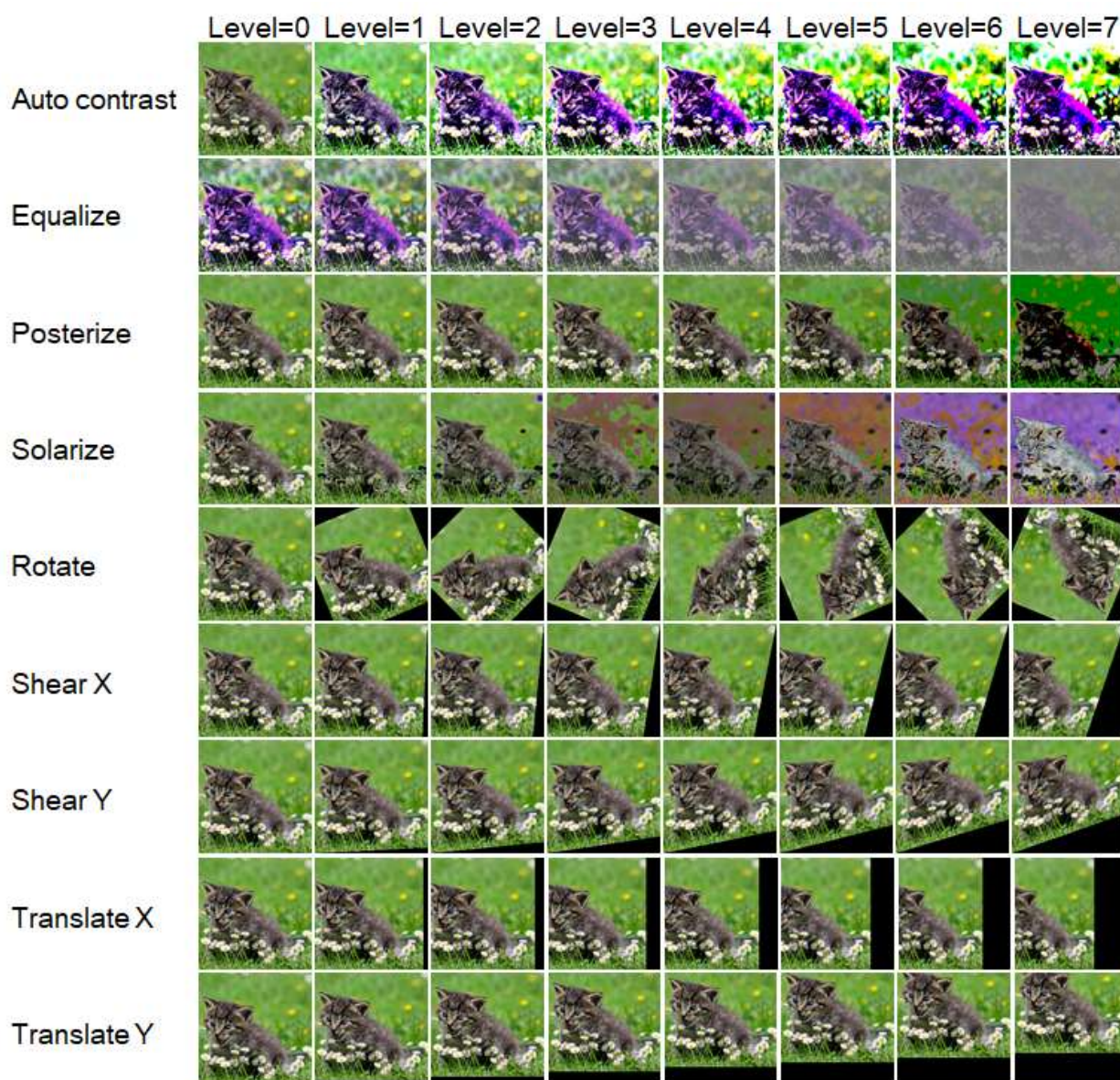


図2 レベル別に各 Augmentation を適用した結果の例
(レベル別の変化を見やすくするため、本来の処理より強調して処理した画像)

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。

2.2 Augmentation

このステップでは原画像に対して 9 種類 (Auto contrast, Equalize, Posterize, Solarize, Rotate, Shear X, Shear Y, Translate X, translate Y) の画像処理 (Operation) をランダムに適用し、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(1)に示す。

$$x_{output}=op_{i,level}(x_{input}),$$

$op \in \{\text{Auto contrast, Equalize, Posterize, Solarize, Rotate, Shear X, Shear Y, Translate X, translate Y}\}$

(1)

ここで、 x_{input} は入力画像、 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、 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、 $op_{3,6}$ は入力画像に対して Posterize をレベル 6 の強度で適用する。その次に各画像に対して上記の処理過程を繰り返し、より多様な拡張画像を生成する。

図2にレベル別に各 Augmentation を適用した結果の例を示す。レベルが大きくなるほど画像の変化量が大きくなる。次に、各 Augmentation について具体的に説明する。

2.2.1. Autocontrast

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 Normalization（正規化）処理である。既存の AugMix では単純な正規化のみを利用したが、提案手法ではレベルの概念を導入し、毎回の Autocontrast 処理で少しずつ異なる画像が生成できるようにする。提案手法では、まず、入力画像のヒストグラムを計算し、設定したカットオフの値（パーセント）を用いてヒストグラム両端の値を 0 に変換する。ヒストグラムのカットオフ式を式(2)に示す。

$$p_{out} = \begin{cases} 0 & \text{if } p_{in} \geq \text{cutoff}_{high} \text{ or } p_{in} < \text{cutoff}_{low} \\ p_{in} & \text{otherwise} \end{cases}, \quad (2)$$

ここで、 p_{in} は処理前の画素値で、 p_{out} は処理後の画素値である。 cutoff_{low} はヒストグラム上で画素を明るい順で並べたときの下位 n パーセントに該当する画素値である。一方、 cutoff_{high} はヒストグラム上で画素を明るい順で並べたときの上位 n パーセントに該当する画素値である。カットオフの基準となる n は式(3)を用いて決める。

$$n = \alpha * \text{level}, \quad (3)$$

ここで、 level は適用する Autocontrast 処理の強度を決めるパラメータである。 α は level のスケールを調節するパラメータで提案手法では 3 に設定する。

次に、ヒストグラムから最も暗い画素を黒 (0) にし、最も明るい画素を白 (255) にするように画像を再マッピングする。ヒストグラム再マッピングの式を式(4)に示す。

$$p_{out} = (p_{in} - \text{Min}) * 255 / (\text{Max} - \text{Min}), \quad (4)$$

ここで、 Min は画像内の最小画素値で、 Max は画像内の最大画素値である。

2.2.2. Equalize

Equalize は画像のヒストグラムを均一化する処理である。既存の AugMix では単純な均一化のみを利用したが、提案手法ではレベルの概念を導入し、毎回の Equalize 処理で少しずつ異なる画像が生成できるようにする。提案手法では、まず、ヒストグラム均一化を行う。ヒストグラム均一化の式を式(5)に示す。

$$H(v) = \text{round}((cdf(v) - cdf_{min}) / (N - cdf_{min}) * 255), \quad (5)$$

ここで、 v は処理前の画素値、 $H(v)$ は処理後の画素値、 $cdf(v)$ はヒストグラム累積関数、 cdf_{min} は最小累積値、 N は総画素数を意味する。次に、レベルによる Equalize 処理の強度を決める式を式(6)に示す。

$$H'(v) = H(v) * (8 - \text{level}) / 8 + 16 * \text{level}, \quad (6)$$

ここで、 $H'(v)$ はレベルによる Equalize 処理を適用する前の画素値で、 $H(v)$ は適用後の画素値である。 level は適用する Equalize 処理の強度である。

2.2.3. Posterize

Posterize は画像の階調 (bit 数) を減らす処理である。既存の AugMix では微小な変動値を利用して毎回異なる強度で Posterize 処理を行うが、その変動の幅が非常に小さく、毎回ほぼ同じような処理になる。これに対して提案手法ではその変動幅を増やしてより多様な処理結果が得られるようにする。提案手法の Posterize 処理式を式(7)に示す。

$$p_{out} = p_{in} \& \sim (2^{\text{level}} - 1), \quad p_{in} \in \{0, 255\}, \quad (7)$$

ここで、 p_{in} は処理前の画素値で、 p_{out} は処理後の画素値である。 level は適用する Posterize 処理の強度を決めるパラメータである。AugMix ではこの値がほぼ 4 に固定されているが、提案手法では 0 から 7 の間の任意の値が選ばれる。 $\sim$ はビット反転 (NOT) 演算子で、 $\&$ は論理積 (AND) 演算子である。例えば、 p_{in} の値が 90 で level は 6 の場合は、 $(2^{\text{level}} - 1)$ は 63 であり、これを 2 進数に変換すると 0011 1111 となる。これにビット反転を行くと 1100 0000 となる。最後にこの値と p_{in} (90 → 0101 1010) の論理積により、1100 0000 & 0101 1010 は 0100 0000 となり、これをまた 10 進数に変換すると 64 となる。このように Posterize は入力画像の階調を 2^8 から $2^{8-\text{level}}$ 段階に減らす処理を行い、上記の例では処理前の各画素値 [0, 255] は処理後 0, 64, 128, 192 の 4 つの値のうちいずれか 1 つに変換される。

2.2.4. Solarize

Solarize は特定条件を満たす画素値を反転する処理である。既存の AugMix では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。これに対して提案手法ではその変動幅を増やしてより多様な処理結果が得られるようにする。提案手法の Solarize 処理式を式(8)に示す。

$$p_{out} = \begin{cases} p_{in} & \text{if } p_{in} < \text{threshold} \\ 255 - p_{in} & \text{otherwise} \end{cases}, \quad (8)$$

ここで、 p_{in} は処理前の画素値で、 p_{out} は処理後の画素値である。 threshold は画素値を反転する基準となる閾値で、入力画像のある画素値がこの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。 threshold の計算式を式(9)に示す。

$$\text{threshold} = 255 * (1 - \beta * \text{level}), \quad (9)$$

ここで、 level は適用する Solarize 処理の強度を決め

るパラメータである．既存の AugMix ではレベルの範囲が 0 から 3 の間であるが，提案手法では 0 から 7 に増やし，より多様な処理結果が得られるようにする． β は *level* のスケールを調節するパラメータで提案手法では 0.1 に設定する．例えば，*level* が 2 である場合，*threshold* は式(9)の計算により 204 となる．画像内に存在するが画素値のうち，204 以上のものに対しては 255 からその画素値を引くことで画素値を反転させる．

2.2.5. Rotate

Rotate は画像を回転する処理である．提案手法の Rotate 処理式を式(10)に示す．

$$\begin{bmatrix} p'_x \\ p'_y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix}, \quad (10)$$

ここで， p_x, p_y は画素 p の処理前の (x, y) 座標で， p'_x, p'_y は画素 p の処理後の (x, y) 座標である． θ は回転する角度 (degree) であり，決め方を式(11)に示す．

$$\theta = \pm \gamma * level, \quad (11)$$

ここで，*level* は適用する Rotate 処理の強度を決めるパラメータである．既存の AugMix ではレベルの範囲が 0 から 3 の間であるが，提案手法では 0 から 7 に拡張した． γ は *level* のスケールを調節するパラメータで提案手法では 3 に設定する．

2.2.6. Shear

Shear は四角形の画像を平行四辺形に変換する処理である．提案手法の Shear 処理式を式(12)に示す．

$$\begin{bmatrix} p'_x \\ p'_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & sh_y & 0 \\ sh_x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix}, \quad (12)$$

ここで， p_x, p_y は画素 p の処理前の (x, y) 座標で， p'_x, p'_y は画素 p の処理後の (x, y) 座標である． sh はせん断因子であり，決め方を式(13)に示す．

$$sh = \pm \delta * level, \quad (13)$$

ここで，*level* は適用する Shear 処理の強度を決めるパラメータである．既存の AugMix ではレベルの範囲が 0 から 3 の間であるが，提案手法では 0 から 7 に拡張した． δ は *level* のスケールを調節するパラメータで提案手法では 0.03 に設定する．Shear X の場合は sh_y が 0 となり，Shear Y の場合は sh_x が 0 となる．

2.2.7. Translate

Translate は画像を平行移動する処理である．提

案手法の Translate 処理式を式(14)に示す．

$$\begin{bmatrix} p'_x \\ p'_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix} * \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix}, \quad (14)$$

ここで， p_x, p_y は画素 p の処理前の (x, y) 座標で， p'_x, p'_y は画素 p の処理後の (x, y) 座標である． t_x, t_y はそれぞれ x 方向， y 方向の移動量であり，決め方を式(15)に示す．

$$t = \pm \epsilon * level, \quad (15)$$

ここで，*level* は適用する Shear 処理の強度を決めるパラメータである．既存の AugMix ではレベルの範囲が 0 から 3 の間であるが，提案手法では 0 から 7 に拡張した． ϵ は *level* のスケールを調節するパラメータで提案手法では 1 に設定する．Translate X の場合は t_y が 0 となり，Translate Y の場合は t_x が 0 となる．

2.3 Mix

このステップでは 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(16)に画像の合成式を示す．

$$x_{mix} = \sum_{i=1}^n w_i * x_i, \quad w_i \in \{0, 1\} \quad (16)$$

ここで， x_i は Augmentation 後の各画像， 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(17)に画像の最終合成式を示す．

$$x_{result} = m * x_{orig} + (1-m) * x_{mix}, \quad m \in \{0, 1\} \quad (17)$$

ここで， x_{orig} は原画像， 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. 実験

3.1 実験環境

学習モデルは WideResnet[9]を，学習データとしては CIFAR-100 および CIFAR-100-C を利用した．2 つとも大きさ 32x32 のカラーデータで 100 クラスで構成されており，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．CIFAR100-C は CIFAR100 データに損傷 (Corruption) を起こしたデータである．Loss 関数は多様な入力範

表 1 各モデル別エラー率の比較結果

	Standard	Cutout[5]	Mixup[6]	CutMix[7]	AugMix[8]	Proposed method
CIFAR-100	22.44	23.39	22.50	22.49	23.30	22.17
CIFAR-100-C	54.04	53.53	50.27	52.87	35.76	35.13

単位[%]

囲にかけて安定的である Jensen-Shannon Divergence Consistency Loss 関数[10]を利用した． バッチサイズは 128, ラーニングレートは 0.1, エポックは 100 と設定し, データ拡張以外は全て同じ設定で実験を行った．

3.2 実験結果

CIFAR-100 および CIFAR-100-C データに対し, Standard (データ拡張を使用せずに WideResnet のみで学習), Cutout, Mixup, CutMix, AugMix そして提案手法を利用した場合のエラー率を比較した． 表 1 にその結果を示す． 提案手法による学習結果は CIFAR-100 の場合, 22.17%であった． これは Standard より 0.27%, Cutout より 1.22%, Mixup より 0.33%, CutMix より 0.32%, AugMix より 1.13%良い結果である． CIFAR-100-C の場合, 提案手法によるテスト結果は 35.13%であった． これは Standard より 18.91%, Cutout より 18.40%, Mixup より 15.14%, CutMix より 17.74%, AugMix より 0.63%良い結果である． 実験結果, 提案手法による学習結果は他の手法と比べて CIFAR-100 および CIFAR-100-C データセット両方で一番良いことが確認された． 特に CIFAR-100 データセットよりノイズが多く入っている CIFAR-100-C データセットで学習の性能向上効果が大きかった．

4. 結論

本研究では, 既存の画像処理と画像合成を利用した手法にランダム適用レベルを加えて, 決定境界付近でより多量の拡張画像を生成し, これらを学習することで既存の手法より高性能なデータ拡張手法を提案した． 実験結果, 既存の手法より性能を向上させることができた．

今後の課題として学習データに対する Operation 手法の選び方や Operation の適用レベルの決め方についての検討が必要である． また, ImageNet などの他のデータあるいは DenseNet などの他の学習モデルを用いた場合でも提案手法が良い性能であるかに

ついて検証が必要である．

参考文献

- [1] V. Vapnik, et al.: “On the Uniform Convergence of Relative Frequencies of Events to their Probabilities,” Theory of Probability and its Applications, pp.264-280, 1971.
- [2] R. Yamashita, et al.: “Convolutional Neural Networks: An Overview and Application in Radiology,” Insights Imaging, Vol. 9, No.4, pp.611-629, 2018.
- [3] O. Chapelle, et al.: “Vicinal Risk Minimization,” Proceedings of the 13th International Conference on Neural Information Processing Systems, pp.395-401, 2000.
- [4] I. Goodfellow, et al.: “Generative Adversarial Nets,” Advances in Neural Information Processing Systems 27, pp. 2672–2680, 2014.
- [5] T. Devries, et al.: “Improved Regularization of Convolutional Neural Networks with Cutout,” arXiv preprint arXiv:1708.04552, 2017.
- [6] H. Zhang, et al.: “Mixup: Beyond Empirical Risk Minimization,” arXiv preprint arXiv:1710.09412, 2017.
- [7] S. Yun, et al.: “Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features,” Proceedings of the IEEE International Conference on Computer Vision, pp. 6023-6032, 2019.
- [8] D. Hendrycks, et al.: “Augmix: A Simple Data Processing Method to Improve Robustness and Uncertainty,” Proceedings of the International Conference on Learning Representations, 2020.
- [9] S. Zagoruyko, et al.: “Wide Residual Networks,” Proceedings of the British Machine Vision Conference, pp. 87.1–87.12, 2016.
- [10] Q. Xie, et al.: “Unsupervised Data Augmentation,” arXiv preprint arXiv:1904.12848, 2019.