Vol.2018-CVIM-213 No.30
2018/9/21

Data Augmentationの良し悪しの検討 -Frechet Inception Distanceに基づく評価方法-

小林　賢一[1,a)]　辻　順平[1,b)]　能登　正人[1,c)]

概要：Deep Learning を用いた学習では大量の画像データが必要であるが，十分なデータの収集が困難な場合がある．この問題に対して人工的に学習データを増やすことで正答率を向上させる Data Augmentation (DA) がある．一方で DA の適用方法によっては精度が低下する場合もあり，DA により生成された画像がどのような影響を与えるか把握する必要がある．そこで本研究では，学習データと検証データ間の距離を用いて DA の良し悪しをはかる手法を提案する．二つの画像集合間の距離を表す指標である Fréchet Inception Distance (FID) が本提案手法に利用できることを明らかにするために，FID と正答率の相関関係を確認する．

キーワード：data augmentation, deep learning

1. はじめに

Deep Learning を用いた画像処理の研究では大量の画像データが必要であるが，十分なデータの収集が困難な場合がある．この問題に対して人工的に学習データを増やすことで正答率を向上させる Data Augmentation (DA) が提案されている．DA によってデータ量を増加させること，またデータに多様性を持たせることが可能である [1] [2]．一般的に DA の設計は経験や知識を元にデータセットに適した DA が設計される．例えば MNIST で有効な DA 手法は Rotation や Scale，Translation，Elastic distortions である [3] [4] [5]．一方で，CIFAR10 や ImageNet のようなデータセットでは，Cropping や Mirroring，Color-shifting，Whitening が選択される [1]．このように，DA の適用方法によっては正答率を低下させてしまう場合もあり，データセットに適した DA を設計するための専門的知識や経験が必要となる．

一般的に Deep Learning を用いた画像処理の研究では，学習用データと検証データの二つのデータを用いて学習・検証を行う．この際，検証用データの集合に類似したデータが学習用データの集合に存在するほど正答率が向上することが考えられる．このような考えの元に，学習用データと検証用データの近さに基づいて，データの良し悪しを評価するというのが本研究のアイデアである（図 1）．我々は学習用データと検証用データの「近さ」をはかるものとして，二つの画像データ集合の距離を表す Fréchet Inception Distance (FID) に着目した [6]．FID によって算出される学習用データと検証用データの距離に従って，距離が小さい場合は正答率が向上する良い学習用データであり，その逆は正答率を低下させる悪い学習用データであるという仮説を立てた．そこで FID の値と正答率の間に相関があれば上記のアイデアに基づくデータの良し悪しの評価が実現できる．

そこで本研究は，画像集合間の距離を FID に基づき算出し，距離と正答率の間の相関分析を行う．

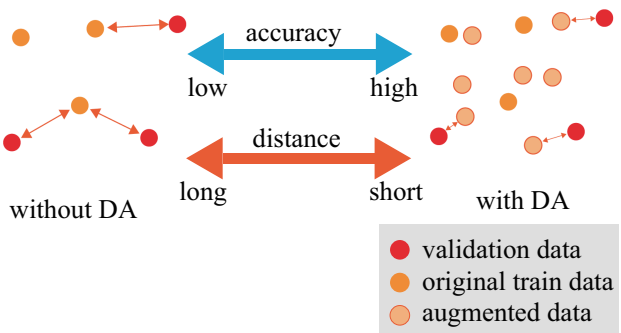


図 1: 本研究のアイデア（距離が短い場合，正答率は向上し，長い場合は低下する．）

[1] 神奈川大学
Kanagawa University
[a)] k-kobayashi@nt.ee.kanagawa-u.ac.jp
[b)] tsuji@kanagawa-u.ac.jp
[c)] noto@kanagawa-u.ac.jp

2. 関連研究

Deep Learning を用いた画像処理を行うためには，大量のデータが必要である．しかしながら，大量データの収集が困難な場合が存在する．そのような少量データの場合において正答率を向上させる方法が複数提案されている [1] [7]．これらの手法を複合させることでより良い精度を得ることが可能である．ここでは，その手法として転移学習と DA の紹介をする．

2.1 転移学習

問題を効果的かつ効率的に解くために，別の関連した問題のデータや学習結果を再利用することを転移学習という [7]．転移する元のデータが，転移先の解きたい問題のデータに類似していれば簡単に転移学習を行うことが可能である．しかしながらそのような場合が難しい問題も存在する．転移学習を用いた具体的な例としては，大規模データセットによって事前学習された学習済ネットワークの識別層を対象とする分類タスクに置き換える．その他の部分は，学習済のパラメータを初期値として用い学習を行うことである．これにより，フルスクラッチから学習するよりもよい結果が得られることが示されている [8]．本研究では，大規模データセット ImageNet をもとに事前に学習したモデルを転移学習したものを用いて実験を行っている．また著者らは，転移学習は DA と同様に正答率を向上させる方法の選択肢の一つであると考えている．

2.2 Data Augmentation

DA とは，既存の学習用データに対し何らかの人工的な加工を施し，学習データを増加させる手法である．具体的には学習データに対して切り出し，平行移動，左右反転，変形，ノイズの付加，RGB 値の操作などが行われている [1] [9] [10] [11] [12]．DA はデータ量を増加させることやデータに多様性を持たせることが可能である [1] [2]．また，DA は設計者の経験や知識を元にデータセットに適した DA を設計される．例えば MNIST で有用な DA 手法は Rotation や Scale, Translation，Elastic distortions である [3] [4] [5]．一方で，CIFAR10 や ImageNet のようなデータセットでは，Cropping や Mirroring や Color-shifting，Whitening が選択される [1]．このように DA を設計するためには，DA の適用方法によっては正答率を低下させてしまう場合もあり，データセットに適した DA を設計するための専門的知識や経験が必要となる．このような背景から DA 手法の選択やパラメータの設定などの戦略を強化学習によって DA を適用する戦略を見つけることも提案されている [13]．

本研究では，データセットに適した DA を設計するための指標として DA により拡張されたデータの良し悪しを評価することが目的であり，その評価指標の検討を行う．

3. FID による画像集合間の距離の評価

FID は，二つの多変量正規分布の平均ベクトルと共分散行列から二つの分布間の距離が計算できる Fréchet Distance [14] を用いて，二つの画像の集合間の距離を求める．この評価指標は，人工的なノイズの変化度合いに対し評価の有用性も示されている．またこの指標は Generative Adversarial Network (GAN) で生成した画像の集合とオリジナル画像の集合の距離を測定し GAN で生成した画像の評価を行う際に用いられている [15]．また，GAN の学習のための指標としても用いられている [16]．具体的な計算方法を下記へ示す．画像の集合を A_i，その要素を $a \in A_i$ とする．ここで Inception-v3 モデルを用いてベクトル $h(a)$ を求める [17]．画像から得られるベクトル $h(a)$ の分布が多変量正規分布に従うと仮定し，二つの多変量正規分布を求めると，その分布間の Fréchet Distance を計算が可能となる．これらを用いて平均ベクトル μ_i と共分散行列 Σ_i を計算する．

$$\mu_i = \frac{1}{|A_i|} \sum_{a \in A_i} h(a) \tag{1}$$

$$\Sigma_i = \frac{1}{|A_i| - 1} \sum_{a \in A_i} (h(a) - \mu_i)(h(a) - \mu_i)^T \tag{2}$$

これらを用いて二つの画像集合 A_1 と A_2 において FID を求める．

$$\mathrm{FID} = |\mu_1 - \mu_2|^2 + Tr(\Sigma_1 + \Sigma_2 - 2(\Sigma_1\Sigma_2))^{\frac{1}{2}} \tag{3}$$

4. FID による DA の良し悪しの評価実験

学習及び検証を行い正答率を算出し，学習データと検証データ間を FID により距離を算出する．これらの正答率と FID との相関関係の分析を行い，距離と正答率に相関関係はあるのかを明らかにする．本研究では，DA 手法を単体で適用する場合と複数の DA 手法を組み合わせて適用する場合の二つの DA 適用方法において実験を行う．

4.1 データセットと CNN モデル

本研究では，MNIST 及び CIFAR10，CIFAR100 の 3 種類のデータセットをそれぞれ用いる．MNIST は 0 から 9 の手書き数字画像（28×28 サイズ）の 10 クラスの構成となっており，合計枚数が 70,000 枚である．CIFAR10 は，10 クラス構成（1 枚が 32×32 サイズ）となっており，合計枚数が 60,000 枚である．CIFAR100 は，100 クラス構成（1 枚が 32×32 サイズ）の合計枚数が 60,000 枚である．本研究の畳み込みニューラルネットワーク（CNN）モデルは全て Deep Learning 用ライブラリ Keras で提供されてい

るモデルを参考に構築している[*1]．MNIST で用いた CNN のモデルの構造は表 1，CIFAR10 で用いた CNN モデルの構造は表 2 に示す．CIFAR100 で用いた CNN モデルの構造は VGG19 である [8]．各データセットは Holdout 検証によりランダムな学習データと検証データの組み合わせを 3 種類（A，B，C）作成する．

表 1: MNIST で用いた CNN の構造

layer	output shape	parameter
Convolution2D	(26, 26, 32)	320
Convolution2D	(24, 24, 32)	9248
MaxPooling2D	(12, 12, 32)	0
Flatten	(4608)	0
Dense	(128)	589952
Dense	(10)	1290

表 2: CIFAR10 で用いた CNN の構造

layer	output shape	parameter
Convolution2D	(32, 32, 32)	896
Convolution2D	(30, 30, 32)	9248
MaxPooling2D	(15, 15, 32)	0
Convolution2D	(15, 15, 64)	18496
Convolution2D	(13, 13, 64)	36928
MaxPooling2D	(6, 6, 64)	0
Flatten	(2304)	0
Dense	(512)	1180160
Dense	(10)	0

4.2 DA 手法とその適用方法

本研究で用いる DA 手法は，上記で説明した 3 種類のデータセットそれぞれで DA を適用した際に正答率の向上が示されている DA 手法を選択している．用いる DA 手法の種類は Rotation，Zoom，Crop，Shear，Cutout の 5 種類である [11] [1]．Rotation はオリジナル画像を元に角度を設定し回転させる手法である．Zoom はオリジナル画像に対して拡大を行う．Crop はオリジナル画像を設定されたサイズに切り抜きを行う．Shear はオリジナル画像の各々の点がある方向へ，その方向と平行な定直線からの符号付き距離に比例して移動するような線型写像である．最後に Cutout は，オリジナル画像をランダムなマスクで欠落させることで，正則化の効果を作り出す手法である．DA の適用は機械学習用 DA ライブラリである Augmentor を用いる[*2]．

また，各 DA 手法の設定値を変換の程度ごとに小さいものから大きいものまでの 5 段階に設定し，それぞれ DA 適用を行う．各 DA 手法の 5 段階の設定値を表 3 に示す（設定値は Augmentor 上のパラメータの設定値である）．DA の適用方法は，DA 手法を単体で適用する実験では，元の学習データに対してデータ量を 2 倍に増加させる．複数の DA 手法を組み合わせて適用する実験では，元の学習データを各 DA 手法を適用しデータ量を 3 倍へ増加させる．

表 3: 各 DA 手法の設定値

	1	2	3	4	5
Rotation (度)	10	15	20	25	30
Crop (pixel)	0.2	0.4	0.6	0.8	1.0
Zoom (pixel)	1.2	1.3	0.6	0.8	1.0
Shear (度)	10	15	20	25	35
Cutout (p)	0.2	0.3	0.4	0.7	0.9

4.3 学習データの良し悪しの評価実験

DA 手法を単体で適用した場合と複数 DA を組み合わせて適用した場合の 2 種類の実験を行う．それぞれの場合で学習及び検証を行い正答率を算出し，学習データと検証データ間を FID により距離を測定を行う．これらの正答率と FID との相関関係の分析を行い，距離と正答率に相関関係はあるのかを明らかにする．

5. 実験結果

5.1 DA 手法を単体で適用した場合

データセット MNIST 及び CIFAR10，CIFAR100 に対し，5 種の DA 手法の 5 段階の設定値をそれぞれ適用した場合の正答率（ACC）と FID を図 2(a) ～ 2(e)（MNIST）及び図 3(a) ～ 3(e)（CIFAR10），図 4(a) ～ 4(e)（CIFAR100）に示す．表 5 に MNIST へ Rotation を適用した際の正答率（ACC）及び FID を示す．また，それぞれの場合の正答率（ACC）と FID の相関係数を表 4 に示す．各 DA 手法の各図及び表の結果より MNIST では Crop, Rotation, Cutout, Shear，CIFAR10 では，Rotation，Shear で，CIFAR100 では Shear で負の相関が見られた．また，FID の距離が大きくなるにつれて正答率も減少していることがわかる．一方で Zoom は他手法よりも相関係数も低く，図からも弱い相関があることがわかる．さらに，表 5 及び各図より，Holdout ごとに DA の程度によらずほぼ同じ FID となっていることが確認できる．

5.2 複数の DA 手法を組み合わせて適用した場合

データセット MNIST, CIFAR10 のそれぞれに対し，Rotation 及び Cutout，Shear を組み合わせて適用した場合の正答率と FID を図 2(f)（MNIST）及び図 3(f)（CIFAR10）

[*1] https://github.com/keras-team/keras-docs-ja
[*2] http://augmentor.readthedocs.io/

表 4: DA 手法を単体で適用した際の相関係数

DA methods	MNIST	CIFAR10	CIFAR100
Without DA	0.9865	0.6048	−0.9765
Crop	−0.8324	−0.4708	−0.4208
Zoom	0.3835	−0.2312	−0.3955
Rotation	−0.8639	−0.8645	−0.1511
Cutout	−0.8174	−0.3877	−0.4021
Shear	−0.6884	−0.6887	−0.6839

表 5: MNIST へ Rotation を適用した際の ACC と FID

Holdouts	DA の程度	ACC(%)	FID
Holdout A	1	99.17	22.12
	2	99.28	22.30
	3	99.24	22.30
	4	99.29	22.29
	5	99.11	22.25
Holdout B	1	98.98	29.22
	2	99.27	29.27
	3	99.15	29.31
	4	99.08	29.40
	5	99.13	29.41
Holdout C	1	98.97	56.78
	2	98.90	56.61
	3	98.64	56.62
	4	98.77	56.61
	5	98.88	56.68

表 6: CIFAR10 へ複数の DA を適用した際の ACC と FID

Holdouts	DA の程度	ACC(%)	FID
Holdout A	1	64.70	54.43
	2	64.37	54.46
	3	62.36	54.65
	4	64.24	54.57
	5	65.57	54.46
Holdout B	1	45.89	119.54
	2	47.18	119.78
	3	47.87	119.95
	4	45.48	119.93
	5	44.42	119.72
Holdout C	1	31.90	219.12
	2	32.35	218.94
	3	31.36	219.00
	4	32.75	218.69
	5	31.10	218.82

に示す．CIFAR10 の正答率（ACC）及び FID を表 6 に示す．また，各条件の正答率と FID の相関係数を表 7 に示す．結果より，MNIST では相関係数が −0.6515 であり強い負の相関がみられる．また CIFAR10 では −0.9344 であり，MNIST より強い負の相関係数がみられる．図 2(f) 及び図 3(f)，表 7 より，DA 手法を単体で適用した場合と同様に，Zoom 以外の DA 手法で Holdout ごとに DA の程度によらずほぼ同じ FID となっていることが確認できる．

表 7: 複数の DA 手法を適用した際の相関係数

DA methods	MNIST	CIFAR10
Rotaion + Cutout + Shear	−0.6515	−0.9344

6. 考察

DA 手法を単体で適用した実験及び，複数の DA 手法を適用した実験の結果から DA の程度に従って FID は変化していないものの，ACC と FID には強い負の相関がみられることが分かった．また全体の結果として，Holdout ごとの FID の変化は明らかであり，FID と正答率の間に負の相関がみられる場合は多くあったが，図 2(a) 〜 4 からもわかる通り，DA の程度の変化に対して FID はほとんど変化が見られなかった．以上から FID は DA の良し悪しを評価することが難しいことが考えられる．ほとんど場合でこのような結果になった原因は，FID を算出する際の CNN が原因で DA の程度の変化を無くしてしまうのではないかと考えられるが，より原因を明らかにする必要がある．

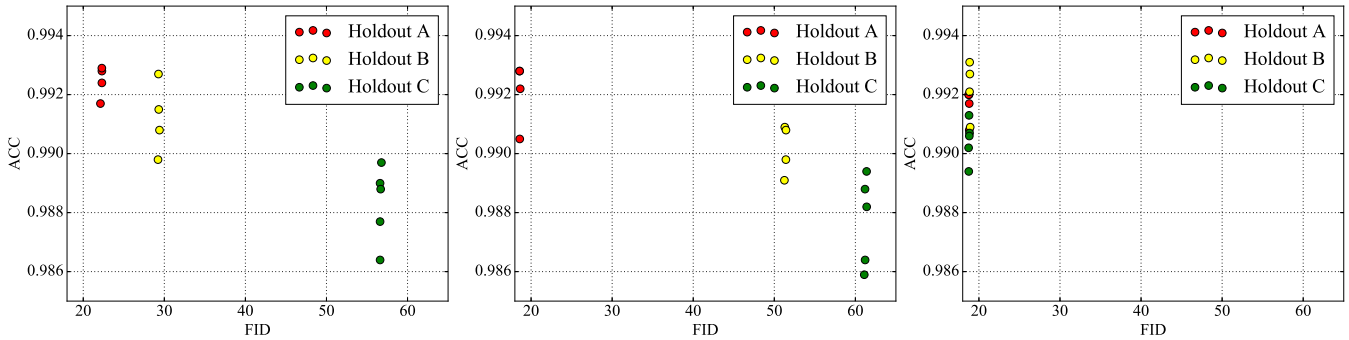
しかしながら，全体の結果として，FID と正答率には負の相関がみられた場合が多々あり，FID で測定した学習データと検証データの距離は，学習データの良し悪しの評価が可能であるのではないかと考えられる．

7. おわりに

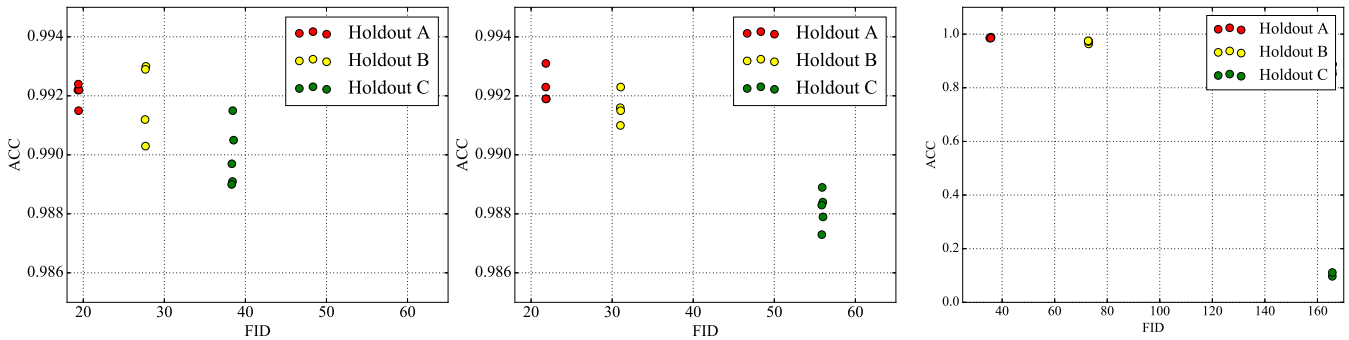
Deep Learning を用いた学習では大量の画像データが必要であるが，十分なデータの収集が困難な場合がある．この問題に対して人工的に学習データを増やすことで正答率を向上させる DA がある．一方で DA の適用方法によっては精度が低下する場合もあり，DA により生成された画像がどのような影響を与えるか把握する必要がある．

そこで本研究では，学習データと検証データ間の距離を用いて DA の良し悪しをはかる手法を提案した．二つの画像集合間の距離を表す指標である FID が本提案手法に利用できることを明らかにするために，FID と正答率の相関関係を確認を行った．

結果は，DA 手法を単体で適用した場合及び複数の DA 手法を組み合わせて適用した場合で，FID と正答率の間に負の相関がみられた場合が多くあった．しかしながら，DA の程度の変化に従い FID はほとんど変化していなく，FID は DA の良し悪しを評価することが難しいと考える．一方，全体の結果として，FID と正答率には負の相関がみられた場合が多々あったことは，FID で測定した学習データと検証データの距離は，学習データの良し悪しの評価が可能であるのではないかと考えられる．

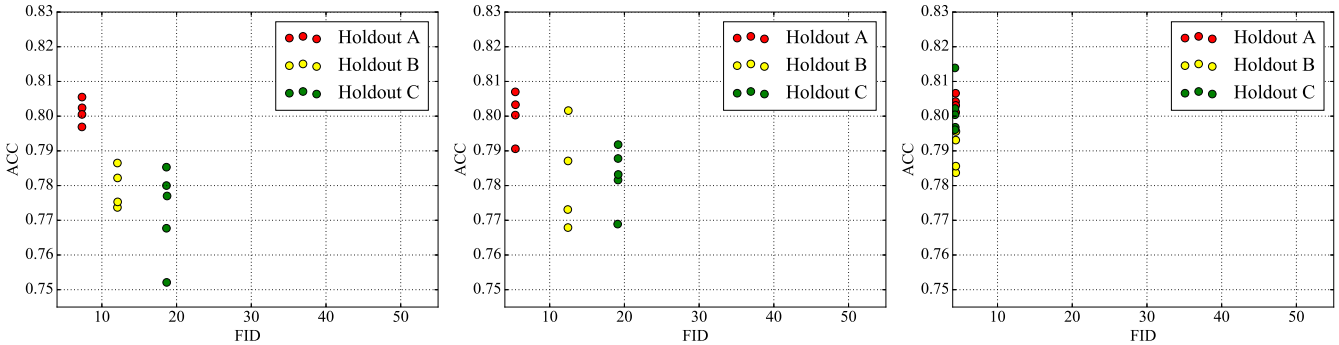


(a) Rotation を適用した場合　(b) Crop を適用した場合　(c) Zoom を適用した場合

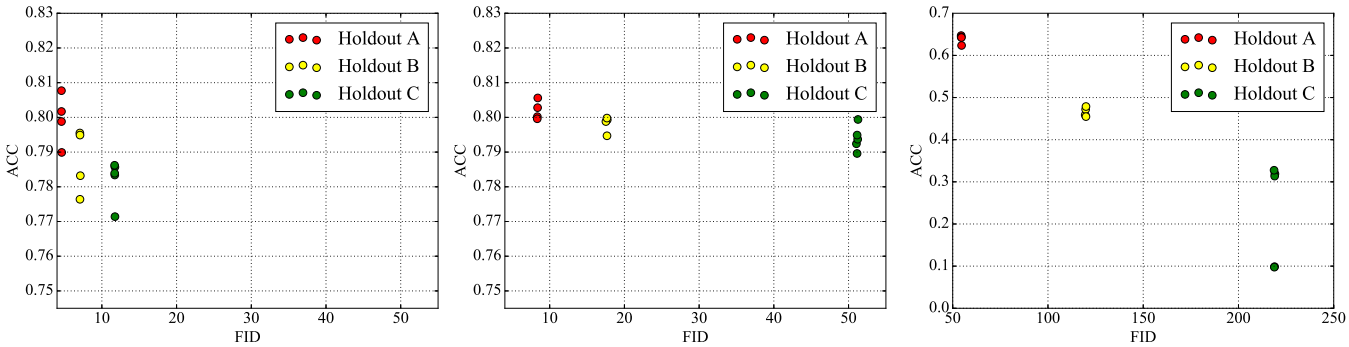


(d) Shear を適用した場合　(e) Cutout を適用した場合　(f) 複数 DA 手法を適用した場合

図 2: MNIST の場合の各 DA 手法の FID 及び正答率（ACC）

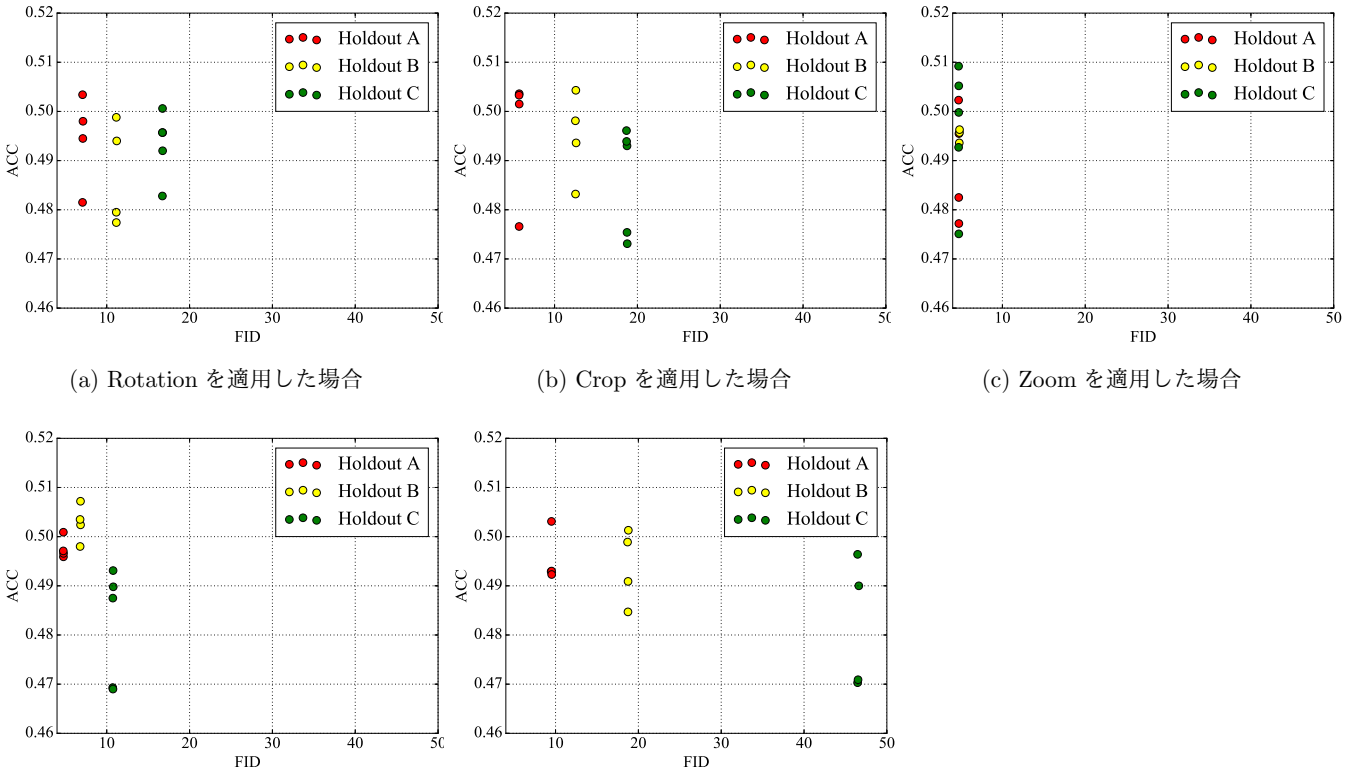


(a) Rotation を適用した場合　(b) Crop を適用した場合　(c) Zoom を適用した場合



(d) Shear を適用した場合　(e) Cutout を適用した場合　(f) 複数 DA 手法を適用した場合

図 3: CIFAR10 の場合の各 DA 手法の FID 及び正答率（ACC）



(a) Rotation を適用した場合　(b) Crop を適用した場合　(c) Zoom を適用した場合

(d) Shear を適用した場合　(e) Cutout を適用した場合

図 4: CIFAR100 場合の各 DA 手法の FID 及び正答率（ACC）

参考文献

[1] Krizhevsky, A., Sutskever, I. and Hinton, G.-E.: ImageNet Classification with Deep Convolutional Neural Networks, *Proceedings of the Advances Neural Information Processing Systems 25*, pp. 1097–1105 (2012).

[2] Simard, P. Y., Steinkraus, D. and Platt, J. C.: Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis, *Proceedings of the Seventh International Conference on Document Analysis and Recognition* (2003).

[3] Dan Ciregan, U. M. and Schmidhuber, J.: Multi-column Deep Neural Networks for Image Classification, *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642–3649 (2012).

[4] Li Wan, M. Z., Zhang, S., LeCun, Y. and Fergus, R.: Regularization of Neural Networks using Dropconnect, *Proceedings of the 30th International Conference on Machine Learning*, pp. 1058–1066 (2013).

[5] Sato, I., Nishimura, H. and Yokoi, K.: APAC: Augmented Pattern Classification with Neural Networks, *CoRR* (2015).

[6] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, *Proceedings of the Advances in Neural Information Processing Systems 30*, pp. 6629–6640 (2017).

[7] 神嶌敏弘：転移学習，人工知能学会誌， Vol. 25, No. 4, pp. 572–580 (2010).

[8] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *CoRR* (2014).

[9] Dosovitskiy, A., Springenberg, J.-T., Riedmiller, M. and Brox, T.: Discriminative Unsupervised Feature Learning with Convolutional Neural Networks, *Proceedings of the Advances in Neural Information Processing Systems 27*, pp. 766–774 (2014).

[10] Gong, Y., Wang, L., Guo, R. and Lazebnik, S.: Multi-scale Orderless Pooling of Deep Convolutional Activation Features, *Proceedings of the 13th European Conference on Computer Vision*, Vol. 8695, pp. 392–407 (2014).

[11] Devries, T. and Taylor, G. W.: Improved Regularization of Convolutional Neural Networks with Cutout, *CoRR* (2017).

[12] Ekin, T., Polson, N. G. and Soyer, R.: Augmented Nested Sampling for Stochastic Programs with Recourse and Endogenous Uncertainty, *Naval Research Logistics*, Vol. 64, pp. 613–627 (2017).

[13] Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V. and Le, Q. V.: AutoAugment: Learning Augmentation Policies from Data, *CoRR* (2018).

[14] Dowson, D.-C. and Landau, B.-V.: The Frechet Distance Between Multivariate Normal Distributions, *Multivariate Analysis*, No. 12, pp. 450–455 (1982).

[15] Lucic, M., Kurach, K., Michalski, M., Gelly, S. and Bousquet, O.: Are GANs Created Equal? A Large-Scale Study, *CoRR* (2017).

[16] Arjovsky, M., Chintala, S. and Bottou, L.: Wasserstein GAN, *CoRR* (2017).

[17] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *CoRR* (2015).