CI-002

DNN-IQA に基づく GAN を用いた高圧縮画像の主観画質改善に関する一検討

A Study on Subjective Image Quality Improvement for High Compressed Images using GAN based on DNN-IQA

和田 直史† 荒澤 孔明† 松﨑 博季† 真田 博文†
Naofumi Wada Komei Arasawa Hiroki Matsuzaki Hirofumi Sanada

岡田 紳太郎‡ 桃井 芳晴‡ 岡崎 敬‡ 山口 寛正‡ 杉村 直純‡
Shintaro Okada Yoshiharu Momonoi Takashi Okazaki Hiromasa Yamaguchi Naozumi Sugimura

1 はじめに

世の中に存在する画像や映像のほとんどは圧縮符号化されたデータである．静止画像では JPEG，動画像では H.265/HEVC などの国際標準符号化方式が一般的に使用されるが，それらの方式は非可逆圧縮であり，復号化後には画質劣化が生じる．圧縮による画質の劣化は，ユーザー満足度を低下させるだけでなく，画像認識に利用する場合には精度低下を引き起こす要因にもなり得る．そのため，圧縮画像の画質改善は重要な課題である．

圧縮符号化による画質劣化（符号化ノイズまたは符号化歪とも呼ぶ）は，見え方の特徴によりブロックノイズやモスキートノイズ，リンギングノイズなどと呼ばれるが，いづれも符号化プロセスにおける “量子化” が発生原因である．JPEG では，離散コサイン変換後の DCT 係数が量子化され，人間の目の知覚感度が低い高周波成分の情報が大きく削減される．したがって，符号化ノイズは，エッジやテクスチャなど空間周波数が高い領域に発生しやすい傾向がある．

画質改善手法の代表的な例としては，デブロッキングフィルタ [1] などの空間ローパスフィルタがある．空間ローパスフィルタは，不自然な高周波ノイズを低減できるが，副作用としてエッジやテクスチャなどもぼやけてしまう．この副作用を抑制する方法として，局所適応フィルタ [2][3] や，畳み込みニューラルネットワーク（以下，CNN）を用いた手法 [4][6][7] が提案されている．しかしながら，これらの手法は基本的に高周波ノイズを低減することを目的としており，テクスチャなどの失われた高周波成分を復元することはできない．

一方，敵対的生成ネットワーク（以下，GAN）[11] を用いて，失われた高周波成分を復元（生成）することにより画質を改善するアプローチがある．GAN を用いた画質改善は，主に超解像の分野で数多く提案されている [12][14][15]．これらの研究では，Perceptual Loss [16] と呼ばれる損失関数に基づき最適化することで，拡大によるボケを防ぎ，主観画質の向上を実現している．しかしながら，GAN により生成した画像は PSNR や SSIM [19] など一般的な客観評価指標で定量評価することが難しく，定性的な主観評価は人手や時間を要するため効率的ではない．

本研究では，狭伝送帯域での利用を想定し，画質劣化が激しい高圧縮画像を対象として，符号化方式に依存しないポスト処理による画質改善技術の開発を目指す．本稿では，超解像分野で得られた知見を基に，GAN を用いた高圧縮画像の主観画質改善手法を提案する（結果を図 1 に示す）．また，GAN で生成した画像の評価方法として，深層ニューラルネットワークベース画質評価指標（以下，DNN-IQA）[23][24] が有効であることを示し，それに基づいて GAN を最適化する．さらに，Encoder-Decoder 構造を持つ U-Net [28] をベースとしたモデルを導入し，実験によりその有効性を示す．

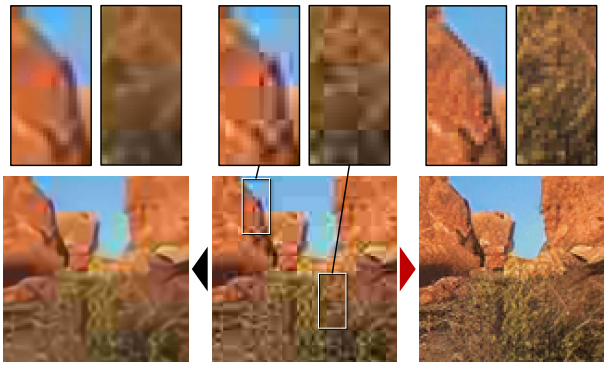

図 1: GAN を用いた高圧縮画像の画質改善の例

2 関連研究

ここでは，本研究に関連する画質改善技術と画質評価指標について述べる．

前述したように，圧縮による画質劣化に対する画質改善手法はこれまでも数多く提案されている．最もシンプルな手法として，デブロッキングフィルタ [1] など符号化ノイズの特徴に合わせた空間ローパスフィルタがあるが，画像をぼかす作用があるためエッジやテクスチャも損なわれてしまう．それに対し，2008 年頃には，SA-DCT [2] や BM3D [3] など，画像の局所構造に合わせて適応的にローパスフィルタを制御し，過度な平滑化を抑制する手法が提案された．2014 年頃からは，画像認識分野で注目を集めた深層学習が，画質改善の分野でも使われるようになった．代表的な手法として，2015 年に Dong らが提案した AR-CNN [4] がある．これは，同著者らが提案した超解像手法である SR-CNN [5] と同じ 3 層から成るネットワークで構成され，深層学習を用いた符号化ノイズ低減手法のベースとなっている．AR-CNN のように，深層学習を用いた符号化ノイズ低減は，超解像と同じフレームワークで実現されることが多い．その後も，CNN を多層化した S-Net [6] や，残差学習（Residual Learning）を導入した DnCNN [7]，Encoder-Decoder 構造を用いた DRUNet [8]，JPEG の品質パラメータの推定と

† 北海道科学大学 工学部 情報工学科
‡ 株式会社サムスン日本研究所

画質改善を同時に行う FBCNN [9] などが提案された．また，近年注目されている Transformer を用いた超解像手法である SwinIR [10] も符号化ノイズ低減が可能である．しかしながら，上記手法は基本的に高周波ノイズを低減することを目的としており，テクスチャなどの失われた高周波成分を復元することはできない．

一方，GAN（Generative Adversarial Networks）[11] を用いて失われた高周波成分を復元（生成）することによって画質を改善する手法が提案されている．GAN は，2014 年に Goodfellow らが提案した画像生成モデルの一つであり，深層学習を用いて画像を生成する生成器（Generator）と真偽判定を行う識別器（Discriminator）を最適化することによって，本物に近い画像を生成する生成器パラメータを獲得する．この GAN を用いた画質改善技術は，主に超解像の分野で数多く提案されている．2017 年に Ledig らが提案した SRGAN [12] は，生成器として ResNet，識別器として VGG [13] を用いた GAN を採用している．また，この SRGAN の改良手法として，生成器に RRDB 構造を導入した ESRGAN [14] や，実際に起こりうる複数の画質劣化（ぼけ，縮小，ノイズ，圧縮）を訓練データに含めて ESRGAN の性能を向上させた Real-ESRGAN [15] が提案されている．上記で挙げた超解像手法は，Perceptual Loss [16] と呼ばれる損失関数に基づいて最適化することで，拡大によるボケを防ぎ，解像感の高い画質を実現している．符号化ノイズ低減においては，Galteri らが SRGAN と同様の ResNet をベースとした GAN による画質改善手法を提案している [17][18]．

一方，上記 GAN を用いた画質改善では，解像感や精細感は向上するものの，それを定量的に評価することが難しいという問題がある．また，定性的な主観評価は人手や時間を要するため，手法を改良するたびに何度も主観評価を行うのは現実的ではない．このように，人間が感じる主観的な画質をコンピュータで計算可能な客観的な評価指標として再現することは非常に困難な問題である．画質評価指標として一般的に使用される PSNR は，原画像との二乗誤差に基づくものであり，わずかなエッジの位置ずれやテクスチャパターンの不一致により評価値が大きく低下することから，主観と一致しない．また，画像の構造・輝度・コントラストに基づく SSIM [19] も頻繁に用いられるが，PSNR より主観に近い値が得られるものの十分とは言えない．このように，原画像を基準として対象画像との差分から評価値を算出する画質評価指標は Full Reference（FR）型と呼ばれる．一方，原画像を必要としない No Reference（NR）型も存在する．NR 型の画質評価指標である BRISQUE [20] や NIQE [21] は，統計的な画像の自然さである NSS（Natural Scene Static）に基づいており，主観との相関が高い．しかしながら，原画像に対する忠実度を無視しているため，画像の種類や被写体によって評価値にばらつきが生じ，画質改善の基準として利用しづらいという問題がある．そのような中，近年，深層ニューラルネットワークを用いた画質評価指標（DNN-IQA）が多く提案されている [22]．FR 型の DNN-IQA である LPIPS [23] や DISTS [24] では，VGG の中間層から得られる特徴マップを比較することで，主観と相関の高い評価を実現するだけでなく，原画像に対する忠実度の評価も同時に実現している．

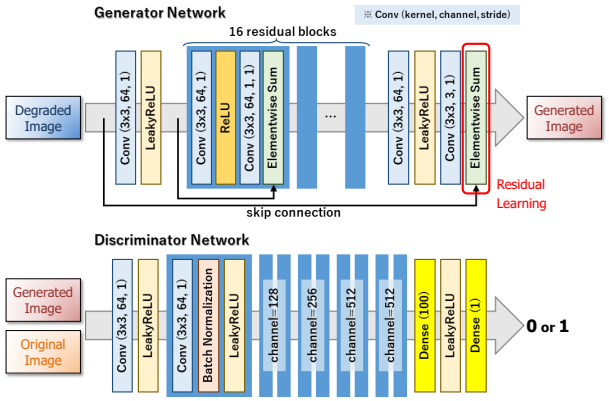

図 2: ResNet をベースとした GAN のモデル

3 GAN を用いた圧縮画像の画質改善

3.1 GAN の学習

ここでは，圧縮前の原画像を I^{HQ}，圧縮後の劣化画像を I^{LQ} とおき，圧縮による劣化過程を次式で表す．

$$I^{LQ} = \mathcal{F}(I^{HQ}) \tag{1}$$

ここで，$\mathcal{F}(\cdot)$ は，JPEG などで圧縮符号化した後，復号化して画像に戻す処理を表す．

本研究では，次式で示すように，劣化画像 I^{LQ} から原画像 I^{HQ} と主観的に同等な高精細画像 $\hat{I}^{HQ}$ を得ることを目的とする．

$$\hat{I}^{HQ} = \mathcal{G}(I^{LQ}) \tag{2}$$

ここで，$\mathcal{G}(\cdot)$ は，劣化画像から高精細画像を復元する処理，すなわち，本研究では GAN で学習した生成器（Generator）による画像生成処理である．GAN の生成器におけるパラメータ $\theta_{\mathcal{G}}$ は次式に基づいて最適化することができる．

$$\hat{\theta}_{\mathcal{G}} = \arg\min_{\theta_{\mathcal{G}}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}(\mathcal{G}_{\theta_{\mathcal{G}}}(I_n^{LQ}), I_n^{HQ}) \tag{3}$$

ここで，$\mathcal{G}(\cdot)_{\theta_{\mathcal{G}}}$ はパラメータ $\theta_{\mathcal{G}}$ を持つ生成器，N は訓練サンプルの数，$\mathcal{L}$ は損失関数，$\hat{\theta}_{\mathcal{G}}$ は最適化後のパラメータである．損失関数 $\mathcal{L}$ については，3.3 で詳しく述べる．

一方，GAN の識別器（Discriminator）は，原画像 I^{HQ} を入力したとき “真”，生成器が出力した画像 $\mathcal{G}(I^{LQ})$ を入力したとき “偽” と正しく判定するように，次式によって最適化する．

$$\begin{aligned} \min_{\theta_{\mathcal{G}}} \max_{\theta_{\mathcal{D}}} & \mathbb{E}_{I^{HQ} \sim p_{train}(I^{HQ})}[\log \mathcal{D}_{\theta_{\mathcal{D}}}(I^{HQ})] + \\ & \mathbb{E}_{I^{LQ} \sim p_{\mathcal{G}}(I^{LQ})}[\log(1 - \mathcal{D}_{\theta_{\mathcal{D}}}(\mathcal{G}_{\theta_{\mathcal{G}}}(I^{LQ})))] \end{aligned} \tag{4}$$

ここで，$\mathcal{D}(\cdot)_{\theta_{\mathcal{D}}}$ はパラメータ $\theta_{\mathcal{D}}$ を持つ識別器を表す．

以上のように，式 (3)(4) に基づき，実際には生成器と識別器のパラメータを交互に更新することによって GAN の学習を行う．

3.2 GAN のモデル

ここでは，Galteri ら [17] と同様，SRGAN [12] を参考にした ResNet ベースの GAN を設計する．生成器と識別器のモデルを図 2 に示す．図 2 上段に示した生成器

は，3×3 の畳み込み層（図中の Conv），活性化関数の LeakyReLU，16 個の残差ブロックで構成されている．また，1 つの残差ブロックは，3×3 の畳み込み層 2 つと活性化関数の ReLU からなり，最後に入力を加算することで Residual 構造 [29] となっている．SRGAN では残差ブロックの中で BN（Batch Normalization）が使用されているが，ESRGAN [14] では BN が悪影響を及ぼすことが示されており，図 2 の生成器においても BN は使用していない．さらに，図 2 の生成器では，最後に入力画像を加えて全体を Residual 構造とする残差学習 [7] を導入している．これにより，入力画像との差分（画質改善するべき成分）のみを効率良く学習するとともに，入力画像と乖離した結果となることを抑制して学習を安定化させる．これら “BN の削除” および “残差学習” の効果については 5.3 節で詳しく述べる．一方，図 2 下段の識別器は，VGG [13] など画像分類で用いられる一般的な識別モデルのように，3×3 の畳み込み層，BN，LeakyReLU を画像を縮小しながら適用し，最後に全結合層を通して真偽を表す 2 値ベクトルを出力する．

3.3 損失関数

ここでは，式 (3) の損失関数 $\mathcal{L}$ について詳しく述べる．損失関数は，生成器が生成した画像 I^{RQ}（式 (3) の $\mathcal{G}_{\theta_G}(I^{LQ})$）と原画像 I^{HQ} との誤差を出力する関数であり，ニューラルネットワークはこの誤差を小さくするように重みを学習する．損失関数 $\mathcal{L}$ は次式で表される．

$$\mathcal{L} = w_1 \cdot \mathcal{L}_{pix} + w_2 \cdot \mathcal{L}_{per} + w_3 \cdot \mathcal{L}_{adv} \tag{5}$$

ここで，$\mathcal{L}_{pix}$ は pixel loss，$\mathcal{L}_{per}$ は perceptual loss，$\mathcal{L}_{adv}$ は adversarial loss であり，w_1〜w_3 は各 loss に対する重み係数である．pixel loss は原画像に対する忠実度を表しており，原画像からの乖離を抑制する働きを持つ．ここでは，画像の幅を W，高さを H とし，L1 ノルムを用いた次式で表す．

$$\mathcal{L}_{pix} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| I^{HQ}_{x,y} - I^{RQ}_{x,y} \right| \tag{6}$$

perceptual loss は，過度な平滑化を抑制しながらテクスチャなどの高周波成分を生成する効果があり，GAN による画質改善において極めて重要な働きを持つ [16]．ここでは，SRGAN と同様，imagenet で学習済みの VGG19 [13] においてプーリング層 5 回，畳込み層 4 回を通った後の特徴マップを使用する．前記特徴マップを取り出す処理を $\phi(\cdot)$ とおくと，perceptual loss は特徴マップの L1 ノルムを用いて次式で表される．

$$\mathcal{L}_{per} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| \phi(I^{HQ}_{x,y}) - \phi(I^{RQ}_{x,y}) \right| \tag{7}$$

adversarial loss は，生成器で生成した画像を識別器に入れたときの出力（0 または 1）の Binary Cross Entropy であり，次式で表される．

$$\mathcal{L}_{adv} = -\frac{1}{N} \sum_{n=1}^{N} \log \mathcal{D}_{\theta_\mathcal{D}}(I^{RQ}_n) \tag{8}$$

式 (5) の重み係数に関しては，経験的に $w_1 = 1.0 \times 10^{-1}$，$w_2 = 1.0$，$w_3 = 5.0 \times 10^{-3}$ とした．

表 1: 図 1 の画像における GAN 適用前後の評価値変化

	PSNR ↑	SSIM ↑	BRISQUE ↓	LPIPS ↓	DISTS ↓
JPEG (*A*)	22.539	0.602	54.206	0.414	0.266
GAN (*B*)	22.342	0.566	15.991	0.253	0.146
Sub. (*B*-*A*)	**-0.197**	**-0.036**	**-38.215**	**-0.161**	**-0.120**

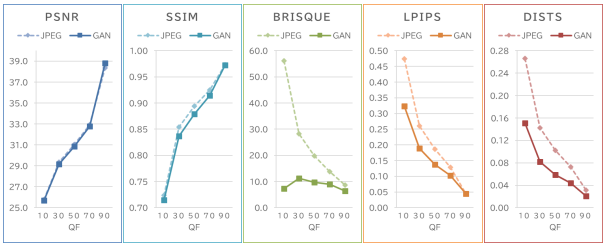

図 3: JPEG 画像の品質が異なる場合の評価値

4 DNN-IQA に基づく GAN

4.1 GAN による画質改善に適した画質評価指標

3.3 節で述べた損失関数に基づいて GAN を最適化することにより，ぼけが少なく見た目に解像感の高い画像を生成することが可能である．しかしながら，原画像を厳密に復元するものではなく，あくまで主観的な画質を向上させるものであるため，PSNR や SSIM ではその効果を正しく評価することができない．そこで，我々は，GAN で生成した画像（以下，GAN 画像）の客観評価に有効な画質評価指標について検討を行った．

ここでは，FR 型の指標として PSNR と SSIM [19]，NR 型の指標として BRISQUE [20]，FR 型の DNN-IQA として LPIPS [23] と DISTS [24] を用いる．LPIPS および DISTS は，どちらも imagenet で学習済みの VGG を通して得られた特徴マップを用いて評価値を算出するが，LPIPS は特徴マップの L2 ノルムを用いるのに対し，DISTS は SSIM と同じように画像構造やコントラストを考慮している点が異なる．PSNR と SSIM は値が大きいほど画質が良く，BRISQUE，LPIPS，DISTS は値が小さいほど画質が良いことを示す．

図 1 中央の JPEG 画像と図 1 右の GAN 画像に対して評価値を算出した結果を表 1 に示す．図 1 では明らかに GAN 画像の主観画質が向上しているにもかかわらず，表 1 の PSNR と SSIM は画質が低下していることを示しており，主観と合致しないことがわかる．一方，BRISQUE，LPIPS，DISTS の 3 つの指標は，画質が向上していることを示しており，主観と合致している．

図 3 は，評価指標ごとに JPEG 画像と GAN 画像の評価値をグラフで表したものである．グラフの横軸は JPEG の Quality Factor（QF）であり，値が小さいほど圧縮率が高く画質劣化が大きくなる．また，実線が JPEG 画像，点線が GAN 画像の結果である．図 3 より，いずれの指標も QF が大きくなるほど JPEG 画像の画質が良くなることを示している．一方，JPEG 画像と GAN 画像の値の差に着目すると，PSNR と SSIM はほぼ差がなく，BRISQUE，LPIPS，DISTS は QF が小さいほど GAN による画質改善効果が高いことを示している．

以上の結果から，GAN の主観画質改善効果を表す指標としては，FR 型の PSNR や SSIM よりも，NR 型の BRISQUE や DNN-IQA の LPIPS および DISTS が有効であることがわかった．

表 2: 損失関数が異なる GAN と非圧縮原画像に対する画質評価値比較

Method	PSNR ↑	SSIM ↑	BRISQUE ↓	VGG54 ↓	LPIPS ↓	DISTS ↓
JPEG (QF=10)	25.754 ±2.161	0.724 ±0.070	56.189 ±10.380	6.462 ±2.788	0.474 ±0.066	0.266 ±0.045
GAN with VGG-loss	25.711 ±2.553	0.715 ±0.091	7.334 ±7.164	**3.948** ±1.931	0.323 ±0.044	0.151 ±0.030
GAN with LPIPS-loss	**26.294** ±2.619	**0.735** ±0.087	11.687 ±8.521	4.021 ±1.990	**0.303** ±0.044	0.148 ±0.029
GAN with DISTS-loss	26.019 ±2.704	0.723 ±0.091	**6.713** ±6.751	4.162 ±2.068	0.321 ±0.044	**0.143** ±0.029
Original (No compression)	inf.	1.000 ±0.000	8.006 ±7.707	0.000 ±0.000	0.000 ±0.000	0.000 ±0.000

4.2 DNN-IQA の損失関数への導入

前節で得られた知見から，DNN-IQA の値を改善するように GAN を最適化することによって，さらなる主観画質改善が可能になると考える．具体的には，3.3 節で述べた損失関数に DNN-IQA を導入する．ここでは，式 (5) における perceptual loss を次の式 (9) または式 (10) のように置き換える．

$$\mathcal{L}_{per} = LPIPS(I^{HQ}, I^{RQ}) \tag{9}$$

$$\mathcal{L}_{per} = DISTS(I^{HQ}, I^{RQ}) \tag{10}$$

ここで，式 (9) を LPIPS-loss，式 (10) を DISTS-loss と呼ぶ．これに対し，従来の式 (7) を VGG-loss と呼ぶ．このとき，LPIPS-loss および DISTS-loss では，VGG-loss の値とスケールを合わせるため，重み係数を $w_2 = 7.0$ とした．

表 2 は，QF=10 の JPEG 画像に対して，VGG-loss，LPIPS-loss，DISTS-loss を用いた GAN の画質評価結果をそれぞれ示している．表 2 では，4.1 節で用いた 5 つの指標の他に，VGG-loss と同じ式 (7) で算出される評価値（VGG54）も指標の一つに加えている．また，表 2 の最下行には，圧縮前の原画像に対して評価値を算出した結果も示している．

まず，原画像に対する評価値に着目すると，FR 型の PSNR，SSIM は最大値を示しているが，NR 型の BRISQUE は値が 0 になっていない．これは，NR 型が原画像に対する忠実度を評価できないことを示している．一方，VGG54，LPIPS，DISTS は特徴マップを比較する FR 型であることから値が 0 になっており，原画像に対する忠実度も評価できていることがわかる．

次に，損失関数が異なる 3 つの GAN に着目すると，それぞれ損失関数として用いている指標の評価値が最も小さい値を示している．このことから，画質評価指標を損失関数に導入することによって，目的の評価値を最適化できていることがわかる．また，LPIPS-loss および DISTS-loss では PSNR と SSIM も向上し，DISTS-loss は BRISQUE の値を最も小さくすることがわかった．

5 評価実験

ここでは，3 章および 4 章で述べた GAN を用いて，高圧縮した JPEG 画像に対する主観画質改善効果を実験により検証する．また，3.2 節で述べた BN の削除と残差学習，および事前学習の効果についても検証する．

5.1 実験条件

モデルの訓練用データとして，超解像評価用のデータセットである DIV2K [25] の訓練用画像 800 枚（画像 1 枚の解像度はおよそ 2K）から，128 × 128 画素の領域をランダムで切り出した 32,208 枚の画像を用いた．また，訓練時には水平・垂直方向の反転および回転によるデータ拡張を適用した．評価用データは，セグメンテーション用のデータセットである BSD300 [26] から評価用の画像 100 枚（画像 1 枚の解像度は 480 × 320 画素）を使用した．また，圧縮符号化方式は JPEG を使用し，圧縮率を決める Quality Factor（QF）を 10, 20, 30, 40 の 4 種類とした（QF が小さいほど圧縮率が高い）．

GAN の学習に関するパラメータは，バッチサイズを 16，イテレーション回数を 100,000（約 50 epoch）とした．最適化手法は Adam [27] を使用し，学習率の初期値を 1.0×10^{-4} とし，学習率はイテレーション回数によって徐々に減衰させた．また，GAN の生成器のみを pixel loss で最適化する事前学習を行い，生成したモデル（重み係数）を GAN の学習の初期値として用いた．GAN の損失関数における perceptual loss には，4.2 節で述べた VGG-loss および DISTS-loss を用いた．

実験に使用した PC は，CPU が Intel Core i9-10980XE，メモリが 256GB，GPU が NVIDIA GeForce RTX3090 × 2 枚である．

5.2 実験結果

実験の結果を表 3 および図 4 に示す．比較手法は，AR-CNN [4]，DRUNet [8]，FBCNN [9]，生成器のみを事前学習した CNN（ResNet-CNN），VGG-loss を用いた GAN（ResNet-GAN with VGG-loss），提案手法である DISTS-loss を用いた GAN（ResNet-GAN with DISTS-loss）の 6 種類である．画質評価指標は，FR 型の PSNR と SSIM [19]，NR 型の BRISQUE [20]，DNN-IQA の LPIPS [23] と DISTS [24] の 5 つの指標を用いた．表 3 の値は，評価用データ 100 枚における RGB の平均値である．

表 3 より，PSNR と SSIM に関しては，DRUNet が最も良い値を示した．一方，主観と相関の高い BRISQUE，LPIPS，DISTS に関しては，GAN を用いた 2 つの手法が最も良い値を示した．特に，DISTS-loss を用いた提案手法は，全ての QF における DISTS で最も良い値を示し，さらに，QF=10，20 では BRISQUE，LPIPS，DISTS でも最も良い値を示した．

図 4 には，TestImage1〜4 に対する各手法の結果画像を示している．図中の (a) は原画像，(b) は QF=10 の JPEG 画像，(c)〜(h) はそれぞれ ARCNN，DRUNet，FBCNN，ResNet-CNN，VGG-loss を用いた GAN，DISTS-loss を用いた GAN の結果を示している．TestImage1 および TestImage2 に関しては，(c)〜(f) の GAN を用いない手法でぼけが生じているのに対し，(g) と (h) の GAN を用いた手法では芝生や腰巻のテクスチャが生成され，主観画質が向上していることがわかる．一方で，TestImage3 に関しては，GAN により解像感は向上しているものの，ややノイジーな結果となった．また，TestImage4 のような幾何学的な構造を持つ人工物では，原画像と同じパターンを復元することが難しいことがわかった．

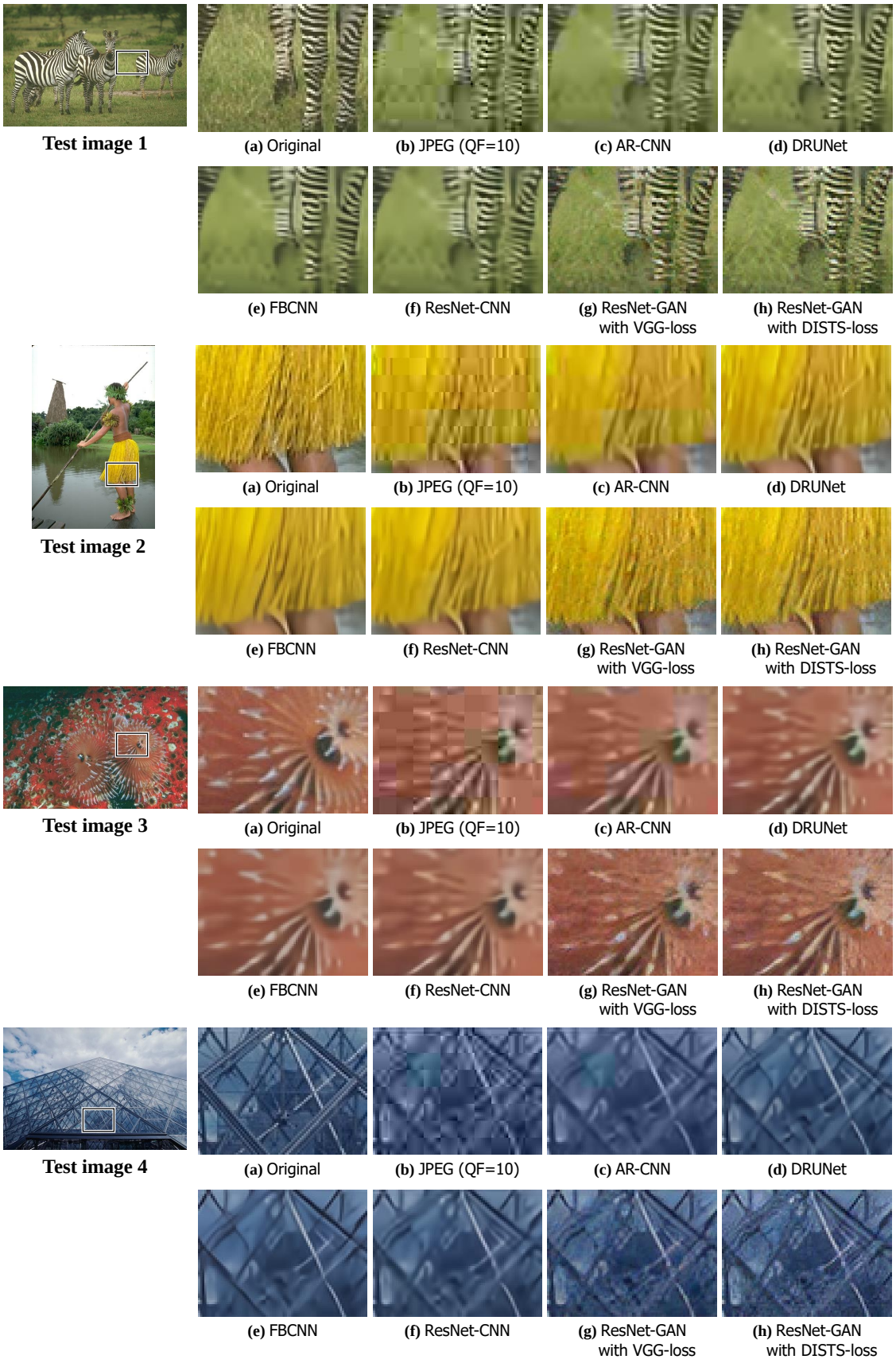

図 4: 主観画質比較

表 3: カラー画像（RGB）における画質評価値の比較（値は RGB の平均値）

QF	Method	PSNR ↑	SSIM ↑	BRISQUE ↓	LPIPS ↓	DISTS ↓
10	JPEG	25.754	0.724	56.189	0.474	0.266
	AR-CNN [4]	26.529	0.745	37.480	0.425	0.230
	DRUNet [8]	27.305	**0.775**	42.389	0.367	0.194
	FBCNN [9]	**27.571**	0.773	42.443	0.368	0.206
	ResNet-CNN	27.298	0.769	40.750	0.386	0.217
	ResNet-GAN with VGG-loss	25.768	0.713	9.128	0.327	0.152
	ResNet-GAN with DISTS-loss (proposed)	25.876	0.714	**8.750**	**0.323**	**0.142**
20	JPEG	28.035	0.814	36.811	0.334	0.181
	AR-CNN [4]	28.821	0.827	31.412	0.311	0.175
	DRUNet [8]	**29.918**	**0.852**	33.401	0.262	0.151
	FBCNN [9]	29.759	0.848	33.206	0.271	0.156
	ResNet-CNN	29.568	0.846	32.491	0.284	0.163
	ResNet-GAN with VGG-loss	27.973	0.798	11.924	**0.232**	0.104
	ResNet-GAN with DISTS-loss (proposed)	28.086	0.800	**10.551**	**0.232**	**0.099**
30	JPEG	29.340	0.854	28.289	0.261	0.142
	AR-CNN [4]	30.164	0.866	25.648	0.251	0.145
	DRUNet [8]	**31.210**	**0.885**	28.240	0.213	0.124
	FBCNN [9]	31.022	0.881	28.228	0.222	0.129
	ResNet-CNN	30.872	0.880	27.713	0.231	0.134
	ResNet-GAN with VGG-loss	29.186	0.839	**10.672**	**0.188**	0.083
	ResNet-GAN with DISTS-loss (proposed)	29.392	0.841	11.364	0.189	**0.079**
40	JPEG	30.267	0.877	23.011	0.217	0.119
	AR-CNN [4]	31.072	0.887	23.664	0.212	0.124
	DRUNet [8]	**32.119**	**0.904**	24.830	0.183	0.107
	FBCNN [9]	31.917	0.901	24.974	0.190	0.112
	ResNet-CNN	31.794	0.899	24.569	0.198	0.116
	ResNet-GAN with VGG-loss	30.061	0.860	**10.320**	0.162	0.069
	ResNet-GAN with DISTS-loss (proposed)	30.304	0.866	10.810	**0.160**	**0.066**

5.3 Ablation Study

ここでは，“事前学習”，“BN の削除”，“残差学習” の 3 つの要素に対して，それぞれの効果を検証した結果を述べる．3.2 節で述べた通り，BN の削除，残差学習は，GAN の学習を安定化するために導入している．また，事前学習は，GAN の生成器のみを pixel loss を最小化するよう最適化し，GAN の学習における重みパラメータの初期値を決める処理である．

各要素を一つずつ無効にした場合の結果を表 4 に示す．表 4 より，いずれの要素を無効にした場合も DISTS の値が増加していることから，“事前学習”，“BN の削除”，“残差学習” の 3 つ要素は全て画質改善に有効であることがわかった．

さらに，事前学習の影響について詳しく述べる．ここでは，事前学習のイテレーション回数によって，その後の GAN の性能がどのように変化するかについて調査した．図 5 のグラフは，横軸が事前学習のイテレーション回数，縦軸が GAN で生成した画像の DISTS である．図 5 より，事前学習で pixel loss に基づいて十分に生成器を最適化しておくことで，その後の GAN の性能が改善することがわかる．特に，イテレーション回数が 0，つまり，事前学習を行わない場合に比べて，事前学習を行った場合の DISTS が大幅に改善していることから，GAN の学習においては，モデルの重みパラメータの初期値を決める事前学習を十分に行うことが有効であることがわかった．

表 4: 事前学習，BN 削除，残差学習の効果比較

	Pre-training	Without BN	Residual Learning	DISTS ↓
Test 1		✓	✓	0.174
Test 2	✓		✓	0.258
Test 3	✓	✓		0.158
Test 4	✓	✓	✓	**0.152**

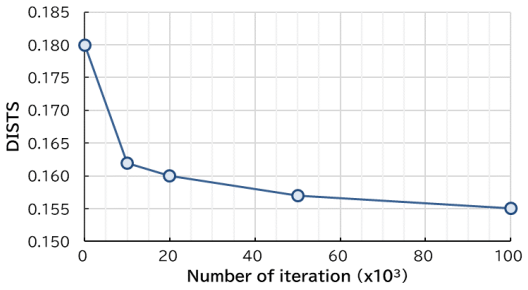

図 5: 事前学習のイテレーション回数による評価値変化

6 U-Net をベースとした GAN モデルの検討

ここでは，Encoder-Decoder 構造を持つ U-Net [28] をベースとしたモデルによる性能改善について述べる．

6.1 生成器への U-Net の導入

3.2 節で述べた ResNet をベースとした生成器では，畳み込みのカーネルサイズが 3×3 と小さいため，特徴マップが考慮する空間的範囲も限定的になる．JPEG では 8×8 画素ブロック，その他の符号化方式では 64×64 画素ブロックを符号化ブロック単位としているものもあり，画質改善を行う際には符号化ブロックよりも広い

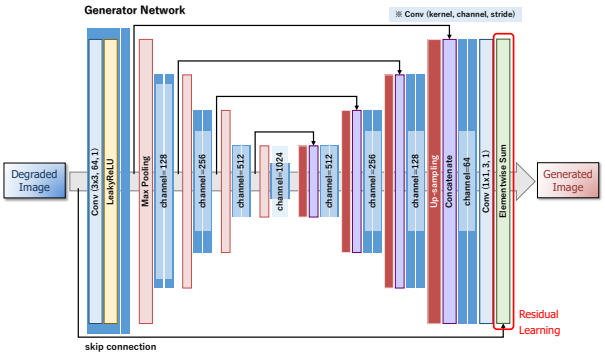

図 6: U-Net（4 層）をベースとした GAN の生成器モデル

範囲の情報から特徴を抽出することが有効であると考える．

そこで，より大局的な特徴を抽出できるよう，Encoder-Decoder 構造を持つ U-Net をベースとしたモデルを生成器に導入する．4 層の U-Net をベースとした生成器を図 6 に示す．ここでは，文献 [28] における通常の U-Net と異なり，図 2 と同様に BN は使用せず，活性化関数には LeakyReLU を使用し，残差学習を導入した．

さらに，U-Net における様々な拡張手法を図 6 のモデルにも導入することが可能である．例えば，U-Net に Residual 構造 [29] を導入した Residual U-Net [30]，Reccurent 構造を導入した R2U-Net [31]，Attention 構造を導入した Attention U-Net [32]，層数の異なる U-Net を Ensemble した U-Net++ [33][34] などがある．

ここでは，図 6 を基本の U-Net とし，前記拡張手法における Residual 構造，Reccurent 構造，Attention 構造，Ensemble 構造を順に導入した，RU-Net，R2U-Net，AR2U-Net，AR2U-Net++を用いて各構造の有効性を評価した．結果を表 5 に示す．表 5 より，Attention 構造以外は DISTS の改善に効果的であることがわかった．また，図 7 は，図 4(h) の ResNet と AR2U-Net++の結果を並べて示したものである．図 7 より，生成器を AR2U-Net++に変更することで，わずかではあるがテクスチャの解像感が向上し，ノイズが低減され，主観画質が改善していることがわかる．

6.2 識別器への U-Net の導入

文献 [35] では，GAN の識別器に U-Net を使用することで局所的な（画素レベルの）評価が可能となり，顔画像生成の精度が向上することが報告されている．また，Real-ESRGAN [15] においても，同様に識別器を U-Net とし，さらに安定化のために SN（Spectral Normalization）[36] を導入している．ここでは，文献 [15] と同様に，識別器に U-Net と SN を導入して性能評価を行った．結果を表 6 に示す．表 6 より，DISTS の値は導入前後で大きくは変わらず，識別器に U-Net を用いることの改善効果はほとんど見られなかった．

7 おわりに

本稿では，非可逆圧縮符号化により劣化した画像に対し，GAN を用いて主観画質を改善する方法ついて検討をおこなった．ここでは，ResNet ベースの生成器を持つ GAN のモデルを提案するとともに，GAN で生成した画像の画質評価において LPIPS や DISTS などの DNN-IQA

表 5: 生成器に U-Net とその拡張手法を導入した結果

Generator	Residual	Reccurent	Attention	Ensemble	DISTS ↓
ResNet (Ref.)	✓				0.142
U-Net					0.137
RU-Net	✓				0.135
R2U-Net	✓	✓			0.133
AR2U-Net	✓	✓	✓		0.136
AR2U-Net++	✓	✓	✓	✓	**0.132**

表 6: 識別器に U-Net を導入した結果

Generator	Discriminator based on U-Net	DISTS ↓
ResNet		0.142
ResNet	✓	0.144
AR2U-Net++		0.132
AR2U-Net++	✓	**0.131**

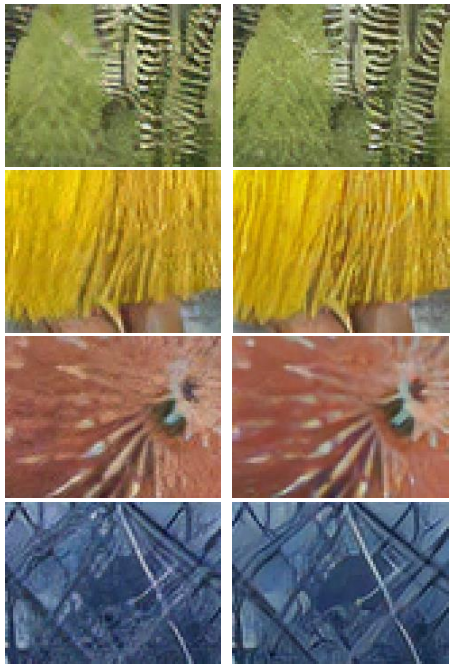

図 7: ResNet と AR2UNet++の主観画質比較

が有効であることを確認した．これにより，DNN-IQA に基づいて GAN の性能検証や性能改善が可能となり，DNN-IQA を損失関数として用いる最適化や，生成器における BN の削除や残差学習，事前学習が有効であることを実験により確認することができた．さらに，生成器を ResNet から U-Net ベースに変更することよって画質評価値が改善することを示した．

今後は，本稿で得られた知見に基づき，より GAN の評価に適した画質評価指標の検討や，それに基づいた更なる画質改善方法の検討を行っていく．

参考文献

[1] T. Jarske, P. Haavisto, and I. Defee, “Post filtering methods for reducing blocking effects from coded images,” IEEE Trans. Consumer Electronics, vol.40, no.3, pp.521-526, Aug. 1994.

[2] A. Foi, V. Katkovnik, and K. Egiazarian, “Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images,” IEEE Trans. Image Processing, vol.16, no.5, pp.1395-1411, May 2007.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” IEEE Trans. Image Processing, vol.16, no.8, pp.2080-2095, Aug. 2007.
[4] C. Dong, Y. Deng, C.C. Loy, and X. Tang, “Compression artifacts reduction by a deep convolutional network,” IEEE International Conference on Computer Vision, Dec. 2015.
[5] C. Dong, C.C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.38, no.2, pp.295-307, Feb. 2016.
[6] B. Zheng, R. Sun, X. Tian, and Y. Chen, “S-Net: A scalable convolutional neural network for JPEG compression artifact reduction,” J. Electronic Imaging, vol.27, no.4, Aug. 2018.
[7] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Processing, vol.26, no.7, pp.3142-3155, July 2017.
[8] K. Zhang, Y. Li, W. Zuo, L. Zhang, L.V. Gool, and R. Timofte, “Plug-and-play image restoration with deep denoiser prior,” IEEE Trans. Pattern Analysis and Machine Intelligence, June 2021.
[9] J. Jiang, K. Zhang, and R. Timofte, “Towards flexible blind JPEG artifacts removal,” IEEE/CVF International Conference on Computer Vision, Sep. 2021.
[10] J. Liang, J. Cao, G. Sun, K. Zhang, L.V.Gool, and R. Timofte, “SwinIR: Image restoration using swin transformer,” IEEE International Conference on Computer Vision Workshops, Oct. 2021.
[11] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Proc. of International Conference on Neural Information Processing Systems, pp.2672-2680, Dec. 2014.
[12] C. Ledig, L. Theis, F. Huszár, J. Caballoero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” IEEE Conference on Computer Vision and Pattern Recognition, July 2017.
[13] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, Sep. 2014.
[14] Z.Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C.C. Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” European Conference on Computer Vision, pp.63-79, Jan. 2018.
[15] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” IEEE International Conference on Computer Vision Workshop, July 2021.
[16] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” European Conference on Computer Vision, Oct. 2016.
[17] L. Galteri, L. Seidenari, M. Bertini, and A.D. Bimbo, “Deep generative adversarial compression artifact removal,” IEEE International Conference on Computer Vision, Oct. 2017.
[18] L. Galteri, L. Seidenari, M. Bertini, and A.D. Bimbo, “Deep universal generative adversarial compression artifact removal,” IEEE Trans. Multimedia, vol.21, no.8, pp.2131-2145, Aug. 2019.
[19] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. on Image Processing, vol.13, no.4, pp.600-612, April 2004.
[20] A. Mittal, A.K. Moorthy and A.C. Bovikm, “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Processing, vol.21, no.12, pp.4695-4708, Dec. 2012.
[21] A. Mittal, R. Soundararajan, and A.C. Bovik, “Making a “Completely Blind” image quality analyzer,” IEEE Signal Processing Letters, vol.20, no.3, pp.209-212, Nov. 2013.
[22] J. Gu, et al., “NTIRE 2021 challenge on perceptual image quality assessment,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, May 2021.
[23] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018.
[24] K. Ding, K. Ma, S. Wang, and E.P. Simoncelli, “Image quality assessment: Unifying structure and texture similarity,” IEEE Trans. Pattern Analysis Machine Intelligence, Dec. 2020.
[25] E. Agustsson, and R. Timofte, “NTIRE 2017 challenge on single image super-resolution: Dataset and study,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 2017.
[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” IEEE International Conference on Computer Vision, July 2001.
[27] D.P. Kingma, and J. Ba, “Adam: A method for stochastic optimization,” arXiv:1412.6980, Dec. 2014.
[28] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” International Conference on Medical Image Computing and Computer-Assisted Intervention, pp.234-241, Nov. 2015.
[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, June 2016.
[30] Z. Zhang , Q. Liu, and Y. Wang, “Road extraction by deep resicual U-Net,” IEEE Geoscience and Remote Sensing Letters, vol.15, no.5, pp.749-753, May 2018.
[31] M.Z. Alom, M. Hasan, C.Yakopcic, T.M. Tahad, and V.K. Asari, “Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation,” arXiv: 1802.06955, Feb. 2018.
[32] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention U-Net: Learning where to look for the pancreas,” Internatinal Conference on Medical Image with Deep Learning, July 2018.
[33] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp.3-11, Sep. 2018.
[34] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,” IEEE Trans. Medical Imaging, vol.39, no.6, pp.1856-1867, June 2020.
[35] E. Schönfeld, B. Schiele, and A. Khoreva, “A U-Net based discriminator for generative adversarial networks,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Aug. 2020.
[36] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida, “Spectral normalization for generative adversarial networks,” International Conference on Learning Representations (ICLR), Feb. 2018.