

2022 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

GAN および Template Matching を用いた

水滴除去手法の研究

Research on Raindrop Removal Method

Using GAN and Template Matching

有本 麻梨香

(5121F005-1)

提出日 : 2023. 01. 23

指導教員 : 渡辺 裕 印

研究指導名 : オーディオビジュアル情報処理研究

# 目次

<b>第 1 章</b>	<b>序論</b> .....	<b>1</b>
1.1	研究の背景 .....	1
1.2	本研究の目的 .....	1
1.3	本論文の構成 .....	1
<b>第 2 章</b>	<b>関連技術</b> .....	<b>3</b>
2.1	まえがき .....	3
2.2	ATTENTIVE GAN .....	3
2.2.1	Attentive GAN とは .....	3
2.2.2	Generator .....	4
2.2.2.1	Attentive-Recurrent Network .....	4
2.2.2.2	Contextual Autoencoder .....	5
2.2.3	Discriminator .....	6
2.3	DEEPFILL v2 .....	7
2.3.1	DeepFill v2 とは .....	7
2.3.2	Generator .....	8
2.3.3	Discriminator .....	8
2.4	テンプレートマッチング .....	8
2.4.1	テンプレートマッチングとは .....	8
2.4.2	SSD .....	8
2.4.3	SAD .....	8
2.4.4	NCC .....	9
2.4.5	ZNCC .....	9
2.5	むすび .....	9
<b>第 3 章</b>	<b>予備実験</b> .....	<b>10</b>
3.1	まえがき .....	10
3.2	データセット .....	10
3.3	ATTENTIVE GAN の再現実験 .....	11
3.3.1	実験 .....	11
3.3.2	考察 .....	11
3.4	水滴の輝度の性質 .....	11
3.4.1	実験 .....	11

3.4.2	考察 .....	13
3.5	CONTEXTUAL AUTOENCODER と DEEPFILL v2 の除去精度の比較.....	13
3.5.1	実験 .....	13
3.5.2	考察 .....	14
3.6	むすび.....	14
<b>第 4 章</b>	<b>提案手法.....</b>	<b>16</b>
4.1	まえがき .....	16
4.2	提案手法の概要 .....	16
4.3	水滴検出 .....	17
4.3.1	概要 .....	17
4.3.2	水滴テンプレート画像の作成 .....	17
4.3.3	テンプレートマッチングの類似度計算方法.....	17
4.4	水滴除去 .....	18
4.5	むすび.....	18
<b>第 5 章</b>	<b>実験と考察 .....</b>	<b>19</b>
5.1	まえがき .....	19
5.2	ATTENTIVE GAN による水滴検出およびマスク生成.....	19
5.3	テンプレートマッチングによる水滴検出およびマスク生成 .....	20
5.3.1	水滴テンプレート画像の作成 .....	20
5.3.2	テンプレートマッチングおよびマスク生成.....	20
5.4	DEEPFILL v2 による水滴除去.....	22
5.5	評価実験 .....	22
5.6	考察 .....	23
5.6.1	提案手法の有効性.....	23
5.6.2	今後の課題 .....	24
5.7	むすび.....	24
<b>第 6 章</b>	<b>結論と今後の展望 .....</b>	<b>25</b>
6.1	結論 .....	25
6.2	今後の展望 .....	25
<b>謝辞 .....</b>		<b>26</b>
<b>参考文献.....</b>		<b>27</b>
<b>図一覧.....</b>		<b>28</b>

表一覽.....	29
研究業績.....	30

# 第1章 序論

## 1.1 研究の背景

近年, 画像の欠落した領域を自動修復する技術として, 画像修復が注目を集めている. 画像修復では, 災害や経年劣化等により一部が欠損した写真を復元したり, 写真に写り込んだ不要物を除去したりすることが可能となる.

画像修復を施す対象物は, 風景や人物など多岐にわたる. その中でもニーズが高まっているものが, 窓やカメラレンズに付着して写真に写り込んだ水滴の除去である. 活用例としては, ドライブレコーダーに付着した雨滴の除去が挙げられる. この技術の進歩により交通事故の減少が期待されているが, 画像処理のアプローチによる水滴除去は未だ実用化に至っていない. その理由は, 従来にも **Attentive GAN** [1]を始めとする様々な水滴除去手法が提案されてきたものの, 撮影時の天候や水滴の背景等により除去対象となる水滴の形状や色は大きく異なり, 完全に水滴を検出・除去してクリアな視界を手に入れるまでの精度を現状上げられていないためである. 車のバックによる交通事故が多発しており, 2022年5月から新型車のバックカメラ装着が義務化される背景からも, 従来と比較して精度の高い水滴除去の実現が急務となっている. その他に, 屋外スポーツの中継カメラ映像や監視カメラ映像においても, 水滴除去の活用が期待されている.

## 1.2 本研究の目的

従来の水滴除去手法としては, ステレオ画像や動画の前後フレームから水滴領域の検出および除去をする **Stereo Waterdrop Removal** [2], **Video-based Water Drop Detection and Removal Method** [3]などが提案されてきた. しかし, ステレオ画像や動画の前後フレームで必ずしも水滴が同じ場所に付着するとは限らない点や, ステレオカメラや動画を用意できるシーンが静止画と比べて限られる点等の問題がある.

そこで, 本研究では単一の静止画に対する水滴除去の精度向上を目指す. 具体的には, 深層学習を用いた水滴除去手法である **Attentive GAN** に対し, 水滴検出および水滴除去のそれぞれを改善し, 水滴が付着していない **Ground Truth** 画像と水滴除去画像間の **PSNR** および **SSIM** 向上を試みる. これにより, 単一の静止画を対象とした水滴除去の精度向上を実現し, ドライブレコーダーやスポーツ中継カメラ, 監視カメラに付着した水滴除去の実用化を目指す.

## 1.3 本論文の構成

本論文の構成を以下に示す.

第1章は本章であり, 本研究における背景, 目的について述べる.

第2章では、本研究で使用する関連技術について述べる。

第3章では、Attentive GAN の再現実験および水滴の輝度の性質について調べた予備実験、Contextual Autoencoder と DeepFill v2 の除去精度を比較した予備実験について述べる。

第4章では、本研究で提案する水滴除去手法について述べる。

第5章では、提案手法を評価する実験について述べ、提案手法の有効性について述べる。

第6章では、本研究における結論と今後の展望について述べる。

## 第2章 関連技術

### 2.1 まえがき

本章では本研究で利用する関連技術として、従来の水滴除去手法である **Attentive GAN**、風景や人物などを対象とした画像修復手法である **DeepFill v2** [4]、入力画像中からテンプレート画像と最も類似する領域を探索するテンプレートマッチングについて述べる。

### 2.2 Attentive GAN

#### 2.2.1 Attentive GAN とは

**Attentive GAN** とは、**Generator** と **Discriminator** から成る **GAN (Generative Adversarial Network)** をベースとした水滴除去モデルである。水滴が付着した単一の静止画が入力されると、**Generator** は水滴領域を検出および除去した画像の生成を試みる。一方で **Discriminator** は、**Generator** が生成した画像が本物に見えるかどうかを検証する。**generative adversarial loss** は以下の式(2.1)で表される。

$$\min_G \max_D \mathbb{E}_{\mathbf{R} \sim p_{\text{clean}}} [\log(D(\mathbf{R}))] + \mathbb{E}_{\mathbf{I} \sim p_{\text{raindrop}}} [\log(1 - D(G(\mathbf{I})))] \quad (2.1)$$

ここで、**G**は **Generator** を、**D**は **Discriminator** を表している。**I**は水滴が付着した画像であり、**Generator** の入力画像である。**R**は水滴が付着していない画像であり、**I**の **Ground Truth** となっている。

また、**Attentive GAN** の **Generator** は水滴領域を検出する **Attentive-Recurrent Network** と、検出した水滴を除去する **Contextual Autoencoder** から成る。図 2.1 は、**Attentive GAN** 全体の構造を示す。

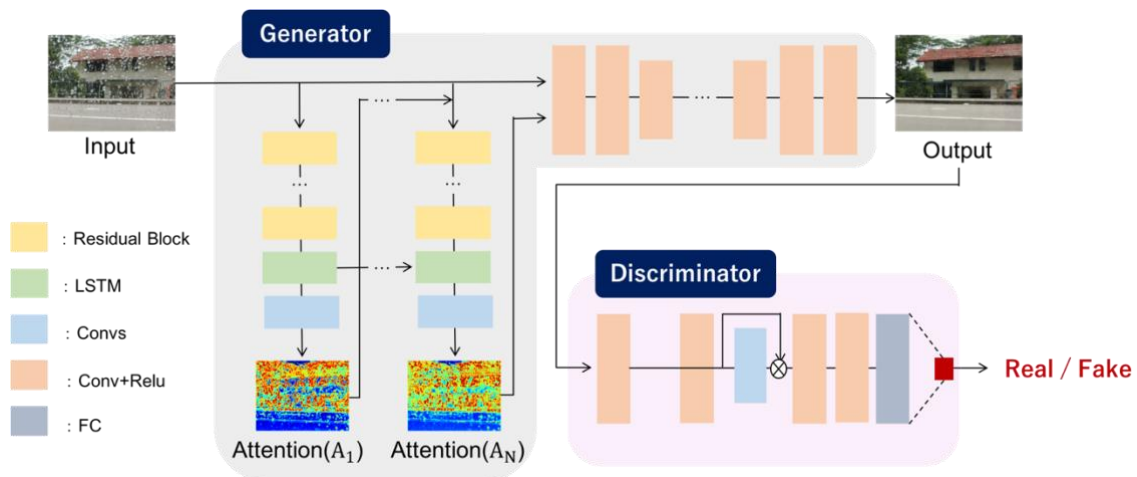


図 2.1 Attentive GAN の構造

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

## 2.2.2 Generator

前述の通り, Attentive GAN の Generator は Attentive-Recurrent Network と Contextual Autoencoder の二つのサブネットワークから成る.

### 2.2.2.1 Attentive-Recurrent Network

Attentive-Recurrent Network は, recurrent network により入力画像から水滴領域を検出し, Attention Map を生成するネットワークである. recurrent network の各ブロックは, 入力画像と前ブロックのマスクから特徴抽出をする 5 層の ResNet[5], LSTM[6], Attention Map を生成する畳み込み層から構成される. Attention Map は 0 から 1 の行列であり, 値が大きいほどその領域が水滴である可能性が高いことを示す. 学習過程では Attention Map の値を 0.5 に初期化する. 各時間ステップでは, 生成された Attention Map と入力画像を連結し, それらを recurrent network の次のブロックに送り込む.

Generator の学習には, 全く同じ背景を含む水滴のある画像と水滴のない画像のペアを使用する. 各 recurrent block の損失関数は, 時間ステップ  $t$  で出力された Attention Map とバイナリマスク間の平均二乗誤差(MSE)で定義される. 損失関数を式(2.2)に示す.

$$\mathcal{L}_{\text{ATT}}(\{\mathbf{A}\}, \mathbf{M}) = \sum_{t=1}^N \theta^{N-t} \mathcal{L}_{\text{MSE}}(\mathbf{A}_t, \mathbf{M}) \quad (2.2)$$

ここで  $\mathbf{A}_t$  は時間ステップ  $t$  で生成された Attention Map,  $\mathbf{M}$  はバイナリマスクを表す.  $\mathbf{A}_t = \text{ATT}_t(\mathbf{F}_{t-1}, \mathbf{H}_{t-1}, \mathbf{C}_{t-1})$  であり,  $\mathbf{F}_{t-1}$  は入力画像と直前の時間ステップの Attention



Map を結合したもの,  $\mathbf{H}_{t-1}$  は直前の時間ステップで LSTM から出力された特徴量,  $\mathbf{C}_{t-1}$  は LSTM の記憶セルである.  $t = 1$  のとき,  $\mathbf{F}_{t-1}$  は 0.5 の値を持つ初期の Attention Map と連結された入力画像である. 関数  $\text{ATT}_t$  は時間ステップ  $t$  における Attentive-Recurrent Network を表す. また,  $N$  は 4,  $\theta$  は 0.8 に設定されている.

### 2.2.2.2 Contextual Autoencoder

Contextual Autoencoder では, Attentive-Recurrent Network で検出された水滴領域を除去し, 水滴のない画像を生成する. Contextual Autoencoder の入力が入力画像と Attentive-Recurrent Network で最終的に生成された Attention Map を連結したものである. 16 個の畳み込みブロックを持ち, 出力がぼやけるのを防ぐためにスキップ接続を追加している. 図 2.2 に Contextual Autoencoder の構造を示す.

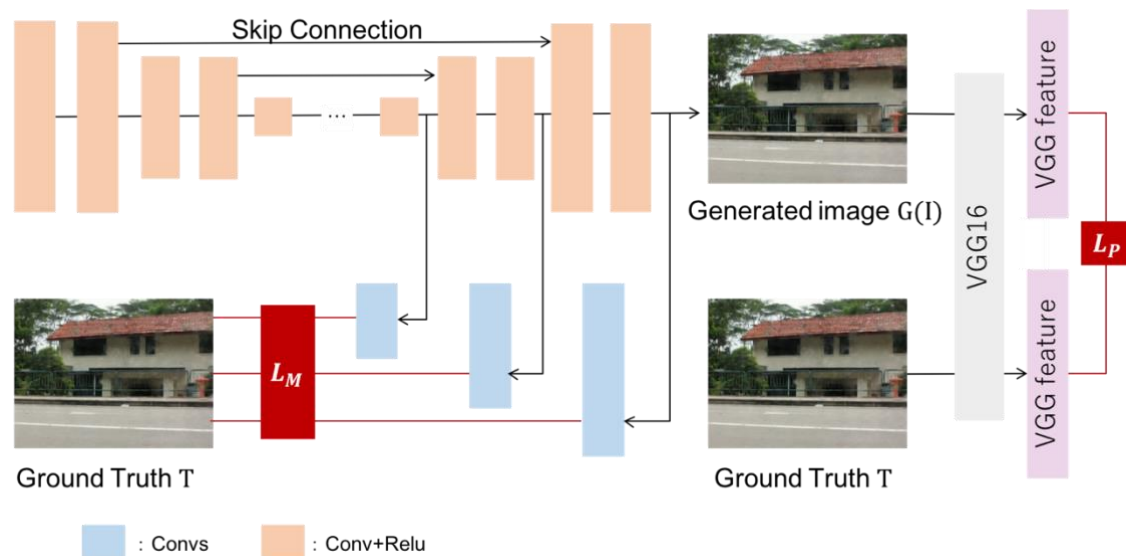


図 2.2 Contextual Autoencoder の構造

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

損失関数は, multi-scale loss と perceptual loss の二つがある. multi-scale loss では異なる decoder 層から特徴を抽出し, 異なるサイズの出力を形成する. これにより, 異なるスケールからより多くのコンテキスト情報を取得することが可能である. multi-scale loss を式(2.3)に示す.

$$\mathcal{L}_M(\{\mathbf{S}\}, \{\mathbf{T}\}) = \sum_{t=1}^M \lambda_t \mathcal{L}_{\text{MSE}}(\mathbf{S}_t, \mathbf{T}_t) \quad (2.3)$$

ここで、 $\mathbf{S}_i$ は decoder 層の*i*番目の出力であり、 $\mathbf{T}_i$ は $\mathbf{S}_i$ と同じスケールの Ground Truth を表す。 $\{\lambda_i\}_{i=1}^M$ は異なるスケールに対する重みであり、スケールが大きいほど重みも大きくなる。

ピクセル単位の操作に基づく multi-scale loss の他に、autoencoder の出力の特徴と対応する Ground Truth 画像の特徴間のグローバルな一致度を測定する perceptual loss がある。これらの特徴は VGG16 から抽出される。perceptual loss を式(2.4)に示す。

$$\mathcal{L}_P(\mathbf{O}, \mathbf{T}) = \mathcal{L}_{\text{MSE}}(\text{VGG}(\mathbf{O}), \text{VGG}(\mathbf{T})) \quad (2.4)$$

VGG は事前学習された CNN であり、与えられた入力画像から特徴量を生成する。 $\mathbf{O}$  は autoencoder の出力画像であり、 $\mathbf{T}$ は水滴のない Ground Truth 画像である。

Generator 全体の損失関数を式(2.5)に示す。

$$\mathcal{L}_G = 10^{-2} \mathcal{L}_{\text{GAN}}(\mathbf{O}) + \mathcal{L}_{\text{ATT}}(\{\mathbf{A}\}, \mathbf{M}) + \mathcal{L}_M(\{\mathbf{S}\}, \{\mathbf{T}\}) + \mathcal{L}_P(\mathbf{O}, \mathbf{T}) \quad (2.5)$$

ここで、 $\mathcal{L}_{\text{GAN}}(\mathbf{O}) = \log(1 - D(\mathbf{O}))$ である。

### 2.2.3 Discriminator

Discriminator は、カーネルが(3,3)の七つの畳み込み層、1024 の全結合層、シグモイド活性化関数を持つ一つのニューロンから成る。内部層から抽出された特徴量を CNN に与え、得られた CNN の出力と Attentive-Recurrent Network から得られた Attention Map に基づいて損失関数が定義されている。これにより、Attention Map で示された領域に焦点を当てて、入力画像が本物か偽物かを判断することができるようになる。Discriminator の損失関数を式(2.6)に示す。

$$\mathcal{L}_D(\mathbf{O}, \mathbf{R}, \mathbf{A}_N) = \log(D(\mathbf{R})) - \log(1 - D(\mathbf{O})) + \gamma \mathcal{L}_{\text{map}}(\mathbf{O}, \mathbf{R}, \mathbf{A}_N) \quad (2.6)$$

ここで、 $\mathcal{L}_{\text{map}}$ は Discriminator の内部層から抽出された特徴量と最終的な Attention Map 間の損失である。式(2.7)に示す。

$$\mathcal{L}_{\text{map}}(\mathbf{O}, \mathbf{R}, \mathbf{A}_N) = \mathcal{L}_{\text{MSE}}(D_{\text{map}}(\mathbf{O}), \mathbf{A}_N) + \mathcal{L}_{\text{MSE}}(D_{\text{map}}(\mathbf{R}), 0) \quad (2.7)$$

ここで、 $D_{\text{map}}$ は Discriminator による 2D マップの生成過程、 $\mathbf{R}$ は水滴のない Ground Truth 画像を表す。  $\gamma$ は 0.05 に設定されている。

## 2.3 DeepFill v2

### 2.3.1 DeepFill v2 とは

DeepFill v2 とは、風景や人物を対象とした GAN ベースの画像修復手法の一つである。 Gated Convolution および SN-PatchGAN のアーキテクチャの採用により、自由形式の領域における画像修復を実現している。 図 2.3 に DeepFill v2 の構造を示す。

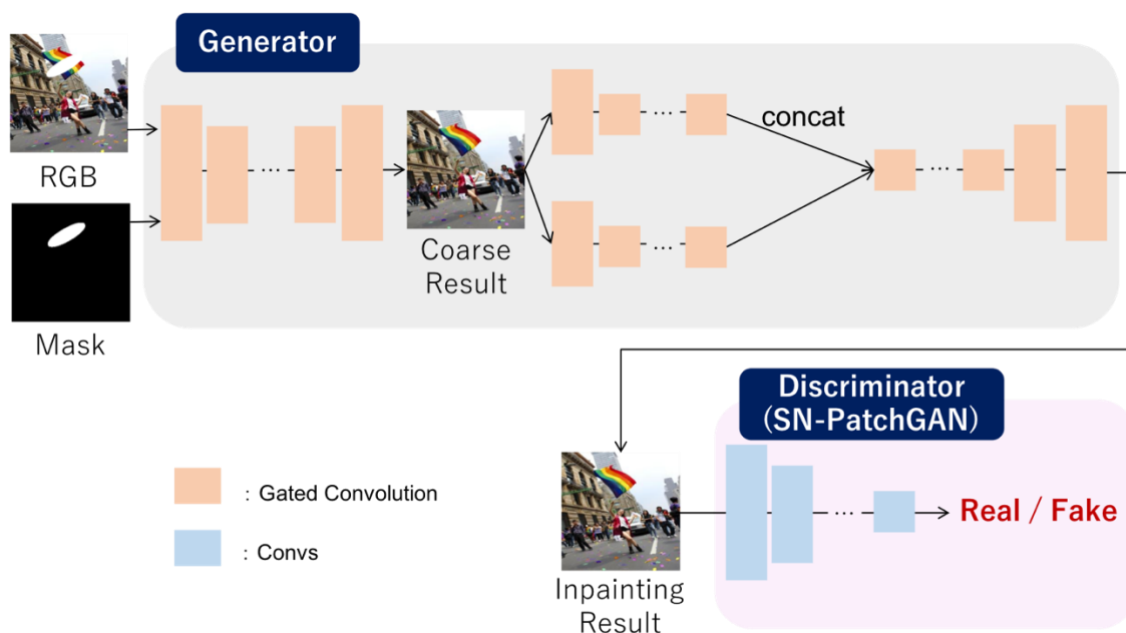


図 2.3 DeepFill v2 の構造

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

損失関数は、adversarial loss のみから成る。損失関数を式(2.8)に示す。

$$\begin{aligned} \mathcal{L}_{D^{\text{sn}}} &= \mathbb{E}_{x \sim p_{\text{data}}(x)} [\text{ReLU}(1 - D^{\text{sn}}(x))] + \mathbb{E}_{z \sim p_z(z)} [\text{ReLU}(1 + D^{\text{sn}}(G(z)))] \\ \mathcal{L}_G &= -\mathbb{E}_{z \sim p_z(z)} [D^{\text{sn}}(G(z))] \end{aligned} \quad (2.8)$$

ここで、 $D^{\text{sn}}$ はスペクトル正則化された Discriminator、 $G$ は Generator、 $z$ は入力画像を表す。

### 2.3.2 Generator

DeepFill v2 の Generator は, マスク領域に対して一度荒く修復するネットワーク(Stage 1)と荒く修復した画像に対して細部を修復するネットワーク(Stage 2)の二つから成る. これらのネットワークはどちらも **Gated Convolution** と呼ばれる畳み込み層から構成される Encoder-Decoder ネットワークである. **Gated Convolution** では学習を通して動的に特徴量を選択し, マスクをアップデートしながら畳み込みを行う. これにより自由形状のマスクに対して, 深層部でもマスク情報を失うことなく学習することが可能となる.

### 2.3.3 Discriminator

Discriminator には, **SN-PatchGAN** が採用されている. **SN-PatchGAN** はカーネルサイズ 5, ストライド幅 2 の六つの畳み込み層から成り, 各畳み込み層から出力された特徴マップに対して, 入力された画像が本物か偽物かの判定を行う. また, 各畳み込み層では, 学習の安定化を目的とした **spectral normalization** を行なっている.

## 2.4 テンプレートマッチング

### 2.4.1 テンプレートマッチングとは

テンプレートマッチングとは, 入力画像中からテンプレート画像と最も類似する領域を探索する処理である. 類似度の計算方法としては, **SSD**, **SAD**, **NCC**, **ZNCC** 等がある.

### 2.4.2 SSD

**SSD** (Sum of Squared Difference)は, 画素値の差分の二乗和である. 値が小さいほど類似度が高いことを表す. **SSD** を式(2.9)に示す.

$$\text{SSD}(d_x, d_y) = \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} (I(d_x + x, d_y + y) - T(x, y))^2 \quad (2.9)$$

ここで,  $w$ はテンプレート画像の幅,  $h$ はテンプレート画像の高さを表す. また, 入力画像の画素値が $I(x, y)$ , テンプレート画像の画素値が $T(x, y)$ , 走査位置が $d_x, d_y$ である.

### 2.4.3 SAD

**SAD** (Sum of Absolute Difference)は, 画素値の差分の絶対値の和である. 値が小さいほど類似度が高いことを表す. **SAD** を式(2.10)に示す.

$$\text{SAD}(d_x, d_y) = \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} |I(d_x + x, d_y + y) - T(x, y)| \quad (2.10)$$

ここで,  $w$ はテンプレート画像の幅,  $h$ はテンプレート画像の高さを表す. また, 入力画像の画素値が $I(x, y)$ , テンプレート画像の画素値が $T(x, y)$ , 走査位置が $d_x, d_y$ である.

#### 2.4.4 NCC

NCC (Normalized Cross Correlation)は, 正規化相互相関である. NCC の値は-1.0~1.0 をとり, 最大値 1.0 に近いほど類似度が高いことを表す. NCC を式(2.11)に示す.

$$\text{NCC}(d_x, d_y) = \frac{\sum \sum [I(d_x + x, d_y + y)T(x, y)]}{\sqrt{\sum \sum [I(d_x + x, d_y + y)]^2} \sqrt{\sum \sum [T(x, y)]^2}} \quad (2.11)$$

ここで,  $\sum \sum = \sum_{x=0}^{w-1} \sum_{y=0}^{h-1}$  である.  $w$ はテンプレート画像の幅,  $h$ はテンプレート画像の高さを表す. また, 入力画像の画素値が $I(x, y)$ , テンプレート画像の画素値が $T(x, y)$ , 走査位置が $d_x, d_y$ である.

#### 2.4.5 ZNCC

ZNCC (Zero Normalized Cross Correlation)は, 零平均正規化相互相関である. ZNCC の値は-1.0~1.0 をとり, 最大値 1.0 に近いほど類似度が高いことを表す. ZNCC を式(2.12)に示す.

$$\text{ZNCC}(d_x, d_y) = \frac{\sum \sum [(I(d_x + x, d_y + y) - \mu_I)(T(x, y) - \mu_T)]}{\sqrt{\sum \sum [I(d_x + x, d_y + y) - \mu_I]^2} \sqrt{\sum \sum [T(x, y) - \mu_T]^2}} \quad (2.12)$$

ここで,  $\sum \sum = \sum_{x=0}^{w-1} \sum_{y=0}^{h-1}$  である.  $\mu_I$ は入力画像の平均値,  $\mu_T$ はテンプレート画像の平均値,  $w$ はテンプレート画像の幅,  $h$ はテンプレート画像の高さを表す. また, 入力画像の画素値が $I(x, y)$ , テンプレート画像の画素値が $T(x, y)$ , 走査位置が $d_x, d_y$ である.

### 2.5 むすび

本章では本研究で用いる Attentive GAN, DeepFill v2, テンプレートマッチングについて述べた.

## 第3章 予備実験

### 3.1 まえがき

本章では, 従来の水滴除去手法である *Attentive GAN* の除去結果を確認する予備実験, 水滴の輝度の性質を調べる予備実験を行う. また, *Attentive GAN* において水滴除去を行う *Contextual Autoencoder* と *DeepFill v2* の同一マスク領域に対する除去精度を比較する予備実験を行う.

### 3.2 データセット

データセットには, 水滴除去の評価に特化した *Attentive GAN dataset* を用いる. そのうち, テストにはテスト用データセットである *test\_a* を用いる. *Attentive GAN dataset* は, 屋外の風景シーンが撮影された水滴なしの画像と同じ風景シーンに水滴が付着している画像のペアが含まれ, 学習用データセットは 1119 ペア, テスト用データセット *test\_a* は 248 ペア用意されている. 水滴が付着している画像は, カメラレンズから 2~5cm 離れた位置に水滴が付着した厚さ 3mm のガラス板を置いて撮影された. 図 3.1 に *Attentive GAN dataset* の画像例を示す.



図 3.1 *Attentive GAN dataset* の画像例

### 3.3 Attentive GAN の再現実験

#### 3.3.1 実験

Attentive GAN の水滴検出および水滴除去精度を確認するため、再現実験を行った。再現実験には、Attentive GAN dataset を 100010 epoch 学習した公開されているモデルを使用した。図 3.2 に test\_a の除去結果の例を示す。

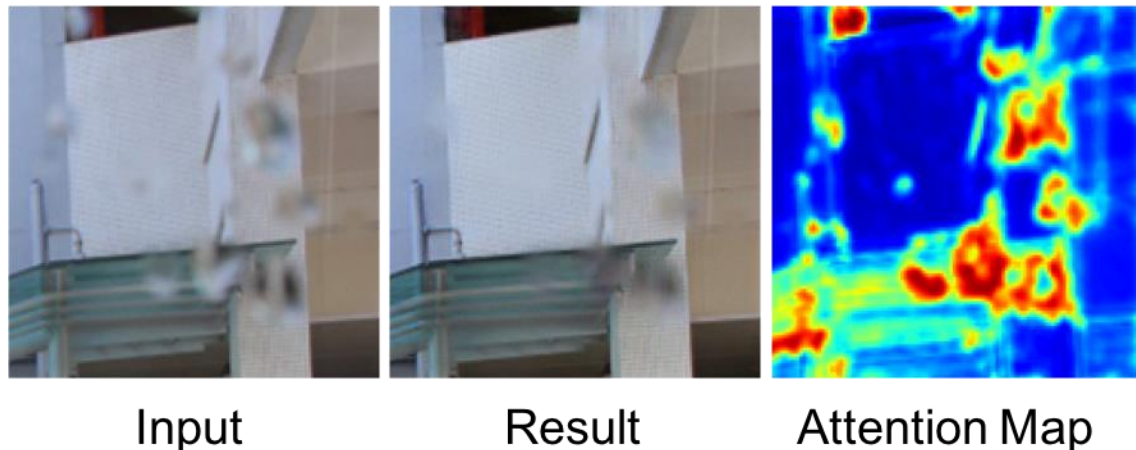


図 3.2 Attentive GAN による水滴除去結果例

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

#### 3.3.2 考察

図 3.2 の水滴除去結果例から、背景シーンの影響を受けて色が一樣でない水滴の除去が完全にできていないことがわかる。これは、水滴の輝度とその時々背景シーンにより変化するため、背景が水滴に反射した領域を水滴として検出することが困難であるためだと考えられる。

### 3.4 水滴の輝度の性質

#### 3.4.1 実験

水滴には背景の逆像が映るため、水滴の  $y$  軸方向の輝度変化と入力画像全体を上下左右反転させた画像の  $y$  軸方向の輝度変化は類似すると考えられる。

そこで、Attentive GAN dataset の画像 3 枚から水滴を 5 つランダムに選択し、水滴の  $y$  軸方向の輝度の平均値と入力画像全体を上下左右反転させた画像の  $y$  軸方向の輝度の平均値を比較した。なお、比較のため入力画像サイズは水滴画像と同じ  $10 \times 10$ [pix] のサイズに縮小している。入力画像と水滴画像はどちらもグレースケール画像である。各画像の比較結果を図 3.3~3.5 に示す。

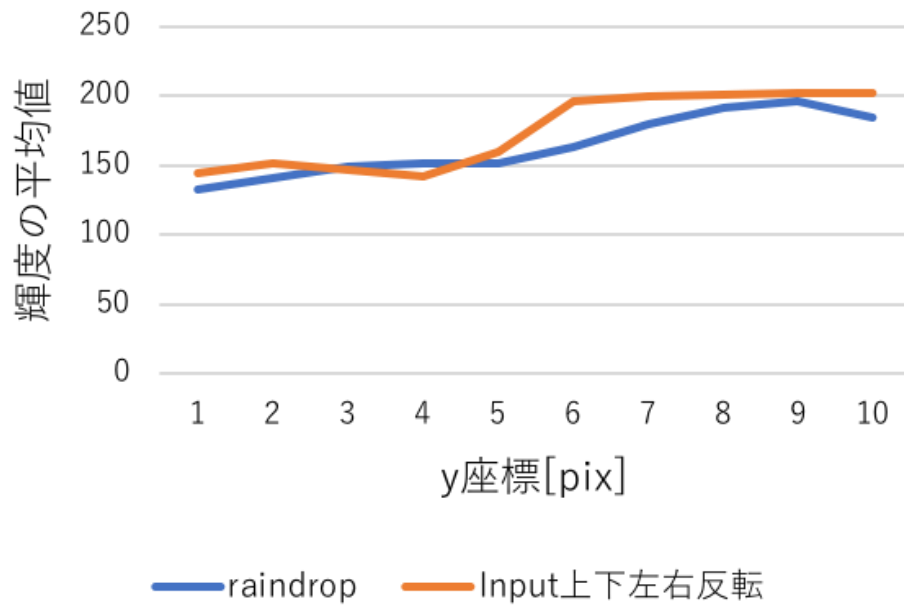


図 3.3 y 軸方向の輝度変化(1)

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

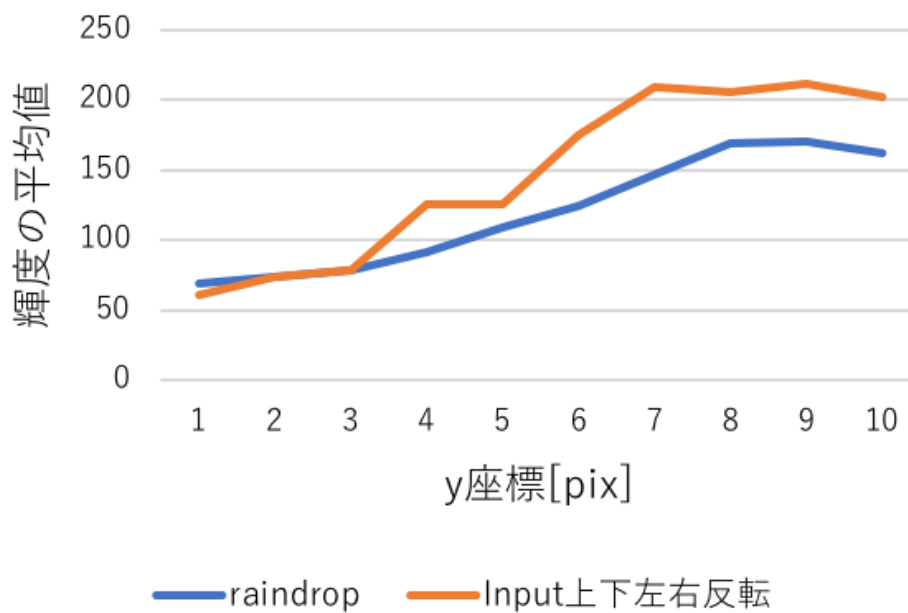


図 3.4 y 軸方向の輝度変化(2)

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)



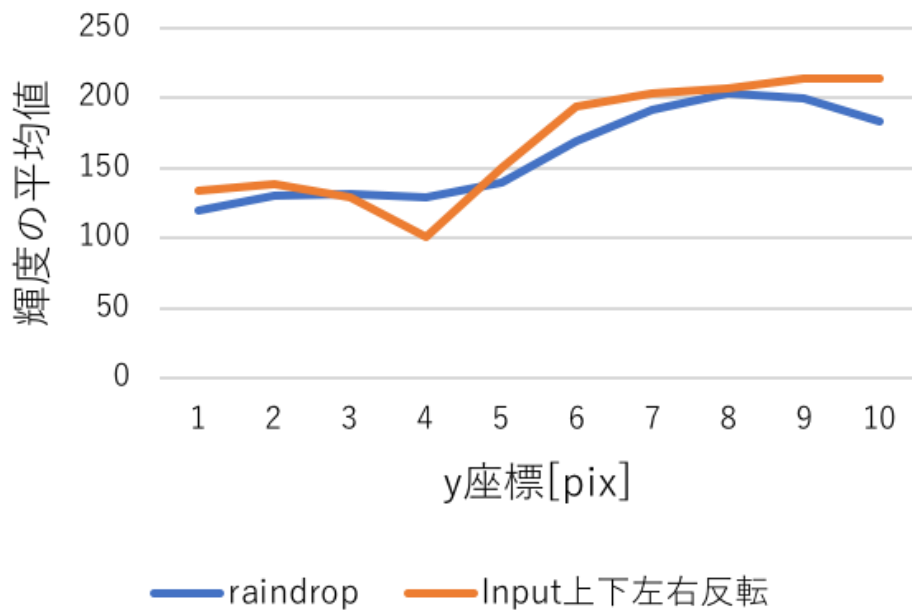


図 3.5 y 軸方向の輝度変化(3)

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

### 3.4.2 考察

図 3.3 から、仮定の通り水滴の y 軸方向の輝度変化と入力画像全体を上下左右反転させた画像の y 軸方向の輝度変化は類似していることがわかる。したがって、Attentive GAN の水滴検出部において、入力画像の逆像の輝度に類似する領域であることを検出条件に加えることにより、水滴検出精度が向上すると考えられる。

## 3.5 Contextual Autoencoder と DeepFill v2 の除去精度の比較

### 3.5.1 実験

同じマスク領域が与えられた場合における、Contextual Autoencoder と Deep Fill v2 の除去精度を比較する実験を行った。Contextual Autoencoder と DeepFill v2 のどちらも、風景シーンデータセットである Places365 [7]を学習したモデルを使用した。実験では、水滴が付着している領域のマスクを手書きで作成し、水滴付着画像とともに Contextual Autoencoder と DeepFill v2 の双方に入力した。図 3.6 に test\_a の水滴除去結果の例を示す。

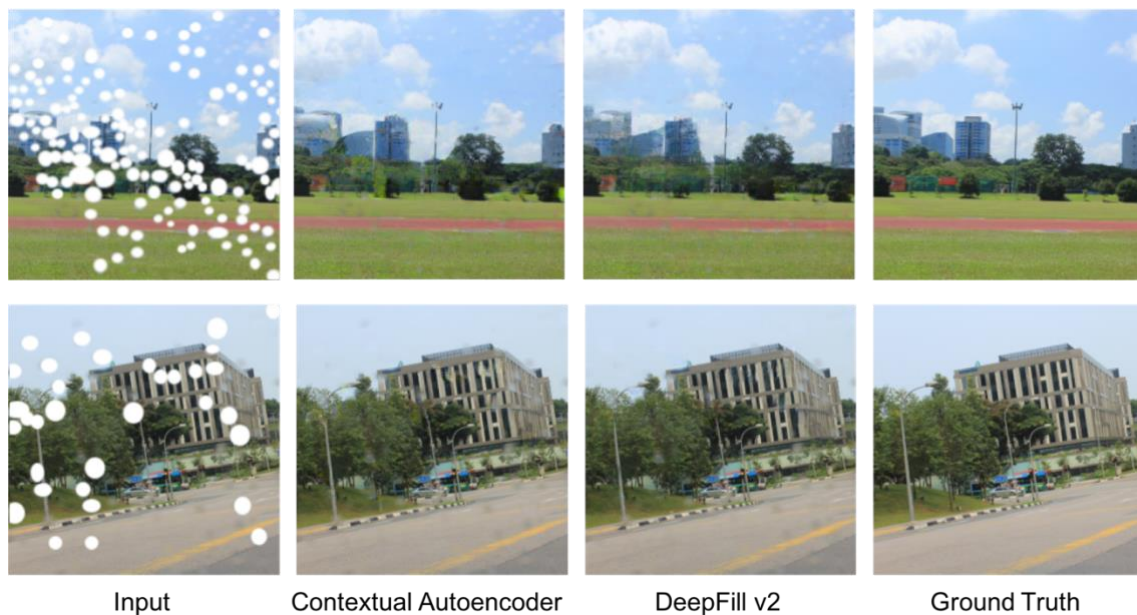


図 3.6 同マスク領域に対する除去結果の比較

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

また，`test_a` からランダムに 5 枚選択し，水滴が付着していない Ground Truth 画像と除去画像の PSNR，SSIM の平均値を算出し，評価を行った．評価結果を表 3.1 に示す．

表 3.1 同一のマスク領域に対する除去結果の PSNR および SSIM (5 枚)

	PSNR(↑)	SSIM(↑)
Contextual Autoencoder	27.1470	0.9409
DeepFill v2	<b>27.4601</b>	<b>0.9426</b>

### 3.5.2 考察

表 3.1 より，同一のマスク領域に対する水滴除去結果において，DeepFill v2 が Contextual Autoencoder より PSNR，SSIM とともに精度を上回る結果となった．したがって，従来の Attentive GAN の水滴除去において用いられていた Contextual Autoencoder から DeepFill v2 に変更することにより，水滴除去の精度向上が期待できる．

## 3.6 むすび

本章では，Attentive GAN の水滴除去結果を確認する再現実験を行った．実験の結果，Attentive GAN では背景シーンの影響を受けて色が一様でない水滴が部分的に除去できていないことがわかった．また，水滴の y 軸方向の輝度変化と入力画像全体を上下反転させた画像の y 軸方向の輝度変化を比較し，それら二つの輝度変化が類似することを確

認した. さらに, 同一のマスク領域に対する水滴除去において, DeepFill v2 が Contextual Autoencoder よりも優れた結果を示すことを確認した.

## 第4章 提案手法

### 4.1 まえがき

本章では, Attentive GAN を改良した水滴除去手法を提案する.

まず, 3.3 節で示した Attentive GAN が背景を反射して輝度変化のある水滴を検出しきれない問題点を解決するため, 背景の逆像を映した水滴を検出するテンプレートマッチングを Attentive GAN の Attentive-Recurrent Network に追加して水滴検出を行う手法を提案する.

また, 3.5 節で前述した通り, 同一のマスク領域に対する除去精度は Contextual Autoencoder より DeepFill v2 が上回ることが確認されたことから, Attentive GAN の Contextual Autoencoder から DeepFill v2 に変更して水滴除去を行う手法を提案する.

### 4.2 提案手法の概要

図 4.1 に提案手法の構造を示す.

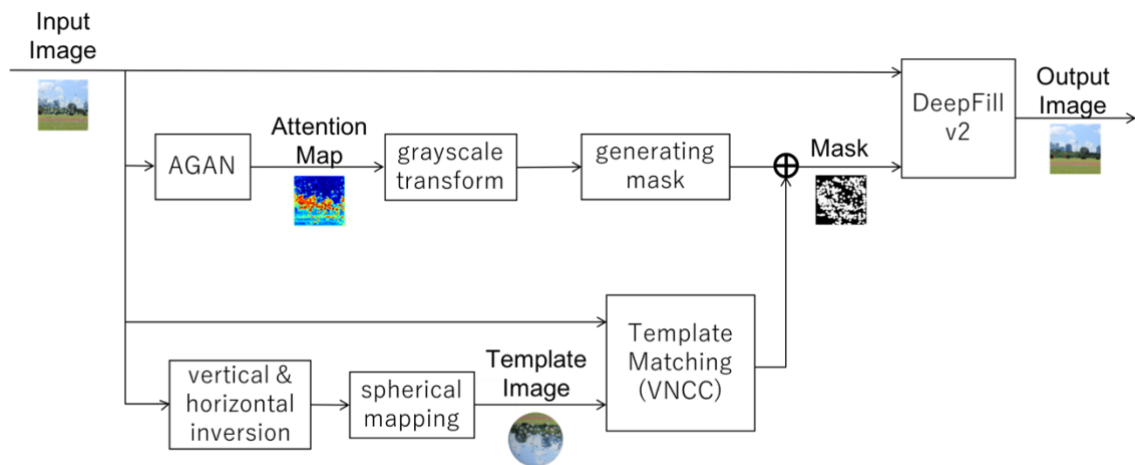


図 4.1 提案手法の構造

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

提案手法は検出した水滴領域のマスクを生成する水滴検出部と, 与えられたマスク領域を除去して周囲の画素値を参照しながら自動的に補完を行う水滴除去部から成り立つ.

## 4.3 水滴検出

### 4.3.1 概要

水滴検出部では，水滴が付着した入力画像を **Attentive GAN** に入力し，**Attention Map** を生成する．**Attention Map** をグレースケール画像に変換し，任意の閾値以上の RGB 値を抽出してマスク画像を生成する．同時に，入力画像を上下左右反転させ，楕円面マッピングを行い，背景の逆像を映した水滴を再現したテンプレートを作成する．入力画像とテンプレート画像のテンプレートマッチングを行い，検出された水滴領域がグレースケール変換された **Attention Map** において任意の閾値以上の RGB 値をとるピクセルを含む場合にのみ，検出された領域が水滴領域であると判断する．テンプレートマッチングの検出領域が **Attention Map** から水滴領域であると判断した場合は **Attention Map** から生成されたマスク画像にマスク領域を追加し，そうでない場合はマスク領域を追加しない処理を行う．これにより，テンプレートマッチングで水滴の誤検出が起きても最終的な水滴検出領域には含まれないようにする．テンプレートマッチングで検出された領域ごとに以上の処理を繰り返すことにより，最終的に水滴領域を示すマスク画像を生成する．

### 4.3.2 水滴テンプレート画像の作成

水滴テンプレートは入力画像を上下左右反転させ，楕円面にマッピングして作成する．マッピングする楕円面の媒介変数表示は極座標系を用いて以下の式(4.1)で求まる．

$$\begin{aligned}
 x &= a \sin \theta \cos \varphi \\
 y &= b \sin \theta \sin \varphi \\
 z &= c \cos \theta \\
 0 \leq \theta \leq \pi, 0 \leq \varphi \leq \pi
 \end{aligned}
 \tag{4.1}$$

一般的に，水滴がガラスに付着した際の接触角は 20~30 度とされる[8]．そこで，接触角 30 度の水滴と同様の形状とするため，マッピングする楕円面は  $a = b = 1, c = 0.2679$  とする．z 軸方向から見た仰角を 0 度，x, y 軸方向の方位角を 90 度にして楕円面を表示した画像を水滴テンプレートとする．

### 4.3.3 テンプレートマッチングの類似度計算方法

入力画像とテンプレート画像間の類似度を計算する方法として **ZNCC** を用いる．これは，**ZNCC** が照明変化にロバストである特徴があり，水滴テンプレートと水滴付着画像間の類似度を計算するには適した方法だと考えられるためである．

#### 4.4 水滴除去

水滴付着画像と水滴検出部で生成されたマスク画像を DeepFill v2 に入力する。DeepFill v2 では、水滴付着画像におけるマスク領域を除去し、除去された領域の画素値を(R, G, B)=(255, 255, 255)に変換する。その後、マスク領域の周囲の画素値を参照しながら自動的にマスク領域の画素を補完する深層学習ネットワークを用いた処理を行う。これにより、水滴を除去した画像を生成する。

#### 4.5 むすび

本章では、Attentive GAN を改良した提案手法の概要について述べた。

3.3 節で示したように Attentive GAN が背景の逆像を映した水滴を検出しきれない問題に対応するため、提案手法では Attentive GAN の Attentive-Recurrent Network で生成された Attention Map に、テンプレートマッチングによる水滴検出結果を加えてマスクを生成する。3.4 節で示したように背景の逆像を映した水滴と入力画像を上下左右反転させた画像の y 軸方向の輝度変化は類似するため、入力画像を上下左右反転させた画像から背景の逆像を映した水滴を再現したテンプレートを作成し、テンプレートマッチングによって検出することで、背景の逆像を映した水滴を完全に検出できると考える。

また、3.5 節で示したように同一のマスク領域に対する除去精度は Contextual Autoencoder より DeepFill v2 が上回ることから、DeepFill v2 により水滴を除去した画像を生成する。これにより、除去単体の精度向上を図る。

## 第5章 実験と考察

### 5.1 まえがき

本章では, 第4章で提案した手法の評価実験とその結果の考察を行う. 提案手法では, Attentive GAN とテンプレートマッチングを用いて検出した水滴領域に対し, DeepFill v2 を用いて水滴除去を行う. データセットは, 3.2節に記載した Attentive GAN dataset test\_a, モデルは Attentive GAN dataset で学習した Attentive GAN の事前学習済みモデル, Places365 で学習した DeepFill v2 の事前学習済みモデルを用いる.

### 5.2 Attentive GAN による水滴検出およびマスク生成

Attentive GAN により水滴が付着した入力画像の水滴領域を検出した Attention Map を生成し, グレースケール画像に変換した. なお RGB の各値は 8bit の整数値をとり, グレースケール画像では RGB の各値が等しい. RGB の各値が任意の閾値以上であるピクセルは  $(R, G, B, A) = (255, 255, 255, 1)$ , 閾値未満であるピクセルは  $(R, G, B, A) = (0, 0, 0, 0)$  の RGBA 画像に変換した. なお RGBA 画像において, 透明度を表すアルファ値は 0~1 をとり, 0 のとき完全に透明, 1 のとき完全に不透明となる.

実験では, 閾値を 90 とした. 図 5.1 に入力画像, Attentive GAN により生成された Attention Map, グレースケール化した Attention Map から生成したマスクを示す.

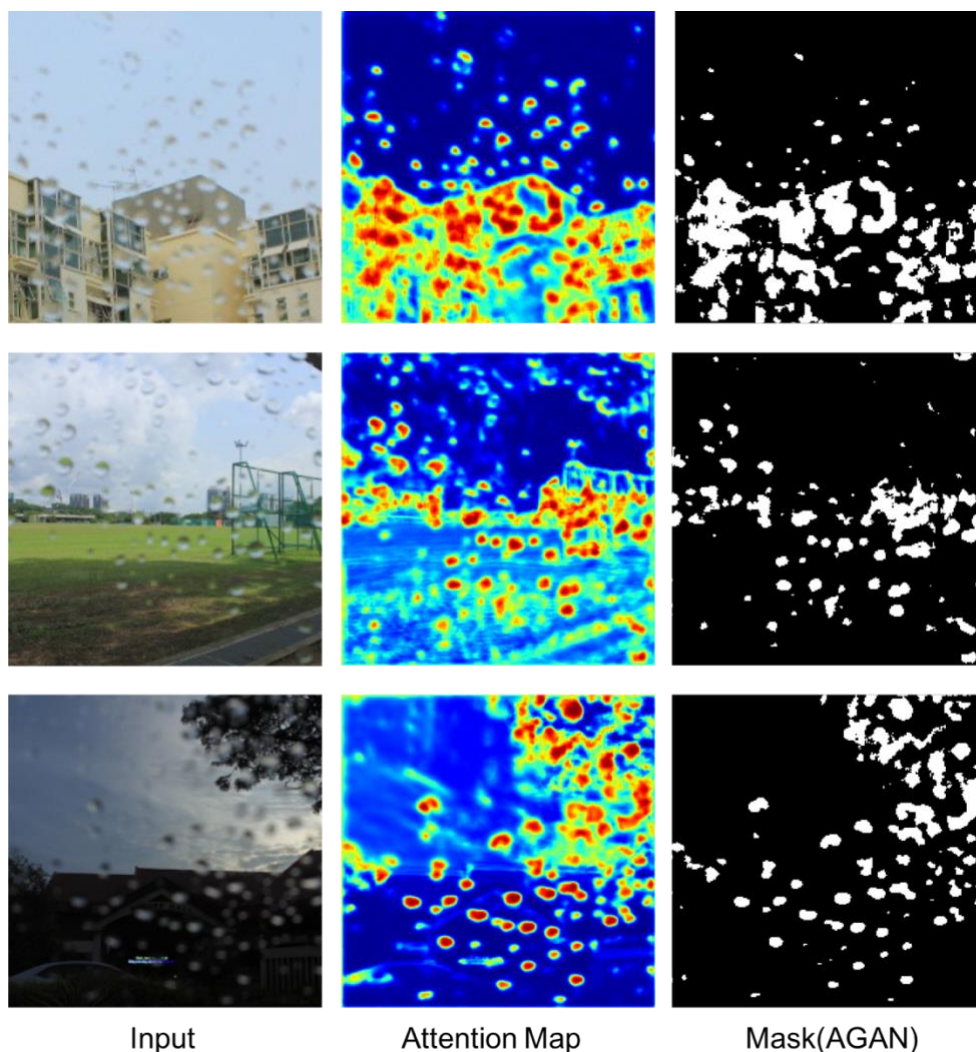


図 5.1 (左から) 入力画像, Attentive GAN により生成された Attention Map, グレースケール化した Attention Map から生成したマスク

### 5.3 テンプレートマッチングによる水滴検出およびマスク生成

#### 5.3.1 水滴テンプレート画像の作成

背景の逆像を映した水滴を再現したテンプレートを作成した. 4.3.2 節で述べたように, 入力画像を上下左右反転させて楕円面にマッピングして得られた画像を水滴テンプレート画像とした. なお水滴テンプレートの画像サイズは  $10 \times 10$ [pix]とし, 水滴以外の領域のピクセルは(R, G, B, A) = (0, 0, 0, 0)の RGBA 画像とした.

#### 5.3.2 テンプレートマッチングおよびマスク生成

入力画像と作成した水滴テンプレート画像のテンプレートマッチングを行った. 実験



では、検出された領域が水滴領域であると判断する ZNCC の閾値を 0.55 とした。また、テンプレートマッチングにより検出された領域が、5.2 節で作成されたマスク画像において  $(R, G, B, A) = (255, 255, 255, 1)$  のピクセルを含む場合にのみ、検出された領域が水滴領域であると判断する。テンプレートマッチングの検出領域が Attention Map から水滴領域であると判断した場合は Attention Map から生成されたマスク画像にマスク領域を追加し、そうでない場合はマスク領域を追加しない処理を行う。

図 5.2 に入力画像、テンプレートマッチングによる水滴検出結果、Attention Map とテンプレートマッチングの水滴検出結果から生成されたマスクを示す。

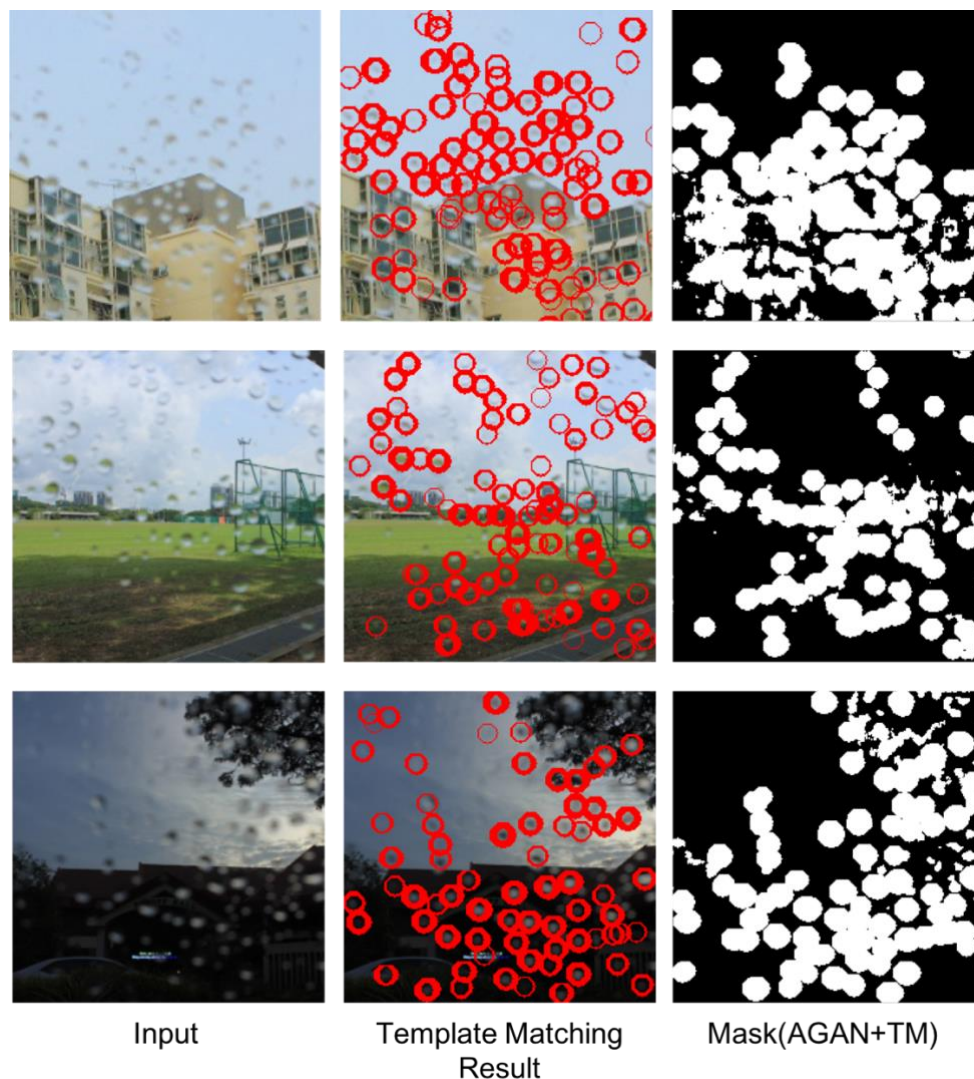


図 5.2 (左から) 入力画像, テンプレートマッチングによる水滴検出結果, Attention Map とテンプレートマッチングにより生成されたマスク画像 (2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

## 5.4 DeepFill v2 による水滴除去

水滴が付着した入力画像と 5.3 節で生成したマスクを DeepFill v2 に入力し、与えられたマスク領域に対して画像修復を行い、水滴除去結果を出力した。また、Attentive GAN から生成された Attention Map にテンプレートマッチングの水滴検出結果を追加することの有効性を確かめるために、5.2 節で示した Attention Map から生成したマスク画像単体を入力画像とともに DeepFill v2 に入力し、水滴除去結果を出力して比較した。図 5.3 に入力画像、Attentive GAN による水滴除去結果、Attentive GAN のみで水滴検出して DeepFill v2 で水滴除去した結果、Attentive GAN およびテンプレートマッチングにより水滴検出して DeepFill v2 で水滴除去した結果（提案手法）、Ground Truth を示す。

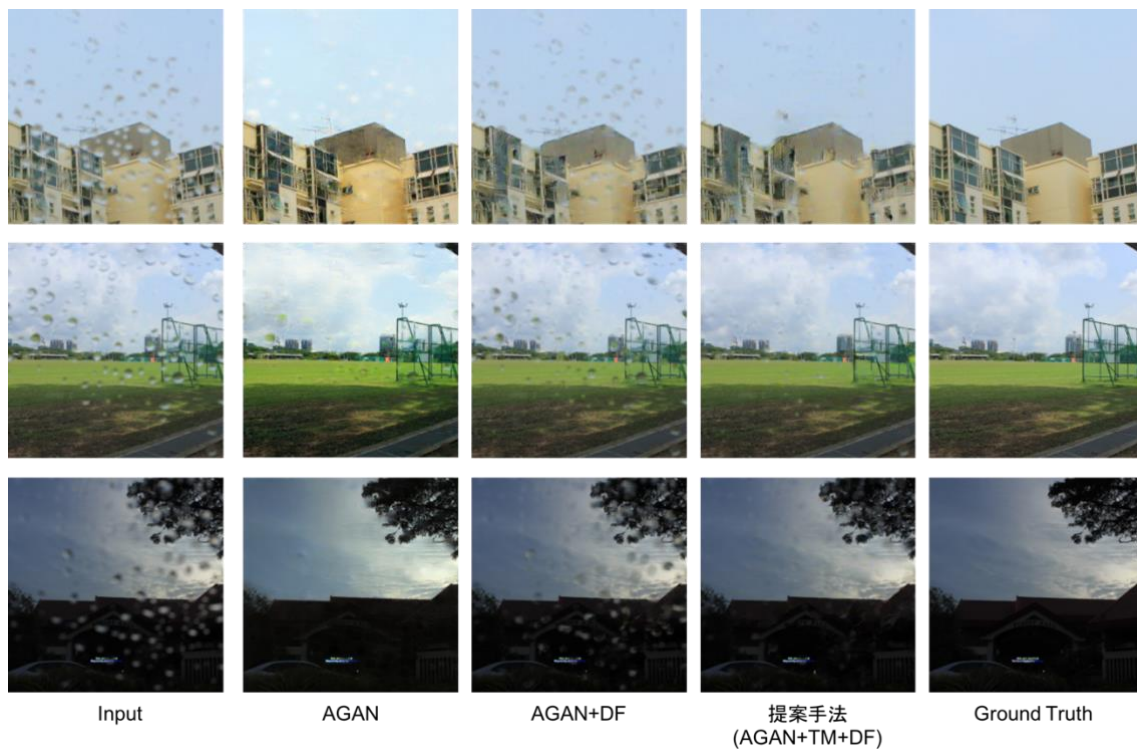


図 5.3 （左から）入力画像，Attentive GAN による水滴除去結果，Attentive GAN および DeepFill v2 による水滴除去結果，提案手法による水滴除去結果，Ground Truth

（2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定）

## 5.5 評価実験

水滴除去結果を評価するため、水滴が付着していない Ground Truth 画像と水滴除去結果間の PSNR, SSIM の平均値をそれぞれ算出した。評価には、test\_a からランダムに 30

枚を選択して得られた各手法の水滴除去結果を用いた。

Attentive GAN による水滴除去結果, Attentive GAN のみで水滴検出して DeepFill v2 で水滴除去した結果, 提案手法により水滴除去した結果と Ground Truth 間の PSNR, SSIM の平均値を表 5.1 に示す。

表 5.1 test\_a の評価結果 (30 枚)

(2023 年情報処理学会オーディオビジュアル複合情報処理研究会にて発表予定)

	PSNR(↑)	SSIM(↑)
AGAN	22.3011	0.8408
AGAN + DF	21.4744	0.8162
提案手法 (AGAN + TM+DF)	<b>23.1711</b>	<b>0.9056</b>

## 5.6 考察

### 5.6.1 提案手法の有効性

表 5.1 に示された結果から, 提案手法は従来手法の Attentive GAN より PSNR, SSIM がどちらも上回ることがわかる。したがって, 提案手法の水滴除去精度は Attentive GAN より高い結果となった。

また, Attentive GAN から生成された Attention Map のみをマスク化して DeepFill で水滴除去した結果は, Attentive GAN より精度が低くなった。Attentive GAN は 0~1 の行列である Attention Map を元に入力画像全域のピクセルに対して RGB 値を再構築して水滴除去結果を出力する。そのため, 水滴の一部がうまく検出できなかった領域も必ず再構築されることとなり, 水滴検出の精度の低さが結果に影響しづらいと考えられる。それに対して Attentive GAN と DeepFill v2 を組み合わせた手法の場合, Deep Fill v2 は二値のマスクで示されたマスク領域に対してのみ再構築を行うため, 水滴が検出されなかった領域に対しては再構築がされず, 一部検出できた領域に対して再構築を行ったとしても, 隣り合うピクセルの影響を受けて水滴の RGB 値を再構築してしまうこととなる。Attentive GAN から出力された Attention Map から閾値以上の輝度を抽出して DeepFill v2 に入力する方法では水滴領域を完全にマスク化することができず, Attentive GAN より精度が低くなったと考えられる。

一方で, 提案手法では Attention Map の結果とテンプレートマッチングの結果を組み合わせることで最終的な水滴検出結果とすることにより, 水滴全域を検出できるようになったと考えられる。3.5 節で示したように DeepFill v2 は Contextual Autoencoder よりも高精度の再構築が可能であるため, Attentive GAN よりも高い精度の水滴除去が可能になったと考えられる。

### 5.6.2 今後の課題

5.6.1 節で述べたように、提案手法では水滴が完全に検出できない場合は DeepFill v2 の Contextual Autoencoder に対する優位性を活かすことができない。

そこで、Attentive GAN の学習用データセットを増やして水滴検出の精度向上を図ることや、DeepFill v2 の入力マスクを二値のマスクではなく 0~1 の行列に変更して、水滴全域が再構築されやすいようにすること等のアプローチにより、提案手法の精度をさらに向上させることができると考えられる。

## 5.7 むすび

本章では、提案手法の有効性を確認するための評価実験の結果を示し、それについての考察と今後の課題を述べた。提案手法は Attentive GAN dataset test\_a の水滴除去において、PSNR と SSIM とともに従来手法の Attentive GAN を上回ることを確認した。今後は学習用データセットを増やしたり、二値のマスクではなく 0~1 の行列で水滴検出結果を出力したりすることで、提案手法の水滴除去精度をさらに向上させることができると考えられる。

## 第6章 結論と今後の展望

### 6.1 結論

本研究では、単一の画像に対する水滴除去手法である Attentive GAN を改良し、水滴除去精度を向上させる提案を行った。

Attentive GAN の問題点としては、第 3 章の予備実験で示したように背景の影響を受けて部分的に水滴の輝度に変化している場合、水滴を検出しきれず除去結果に水滴が残ってしまうことがある。そこで、水滴検出にテンプレートマッチングを追加し、水滴除去を Contextual Autoencoder から DeepFill v2 に変更する方法を提案した。実験により、提案手法は Attentive GAN dataset test\_a の 30 枚に対して PSNR, SSIM とともに向上し、Attentive GAN より高い精度で水滴除去できることを確認した。

### 6.2 今後の展望

従来提案されてきた水滴除去手法は通常のレンズで撮影された画像に対して水滴除去を行うことを前提としており、魚眼レンズで撮影された画像に特化して水滴除去を行う手法はまだ提案されていない。しかし、通常レンズより広い視野で撮影できる魚眼レンズはドライブレコーダーや監視カメラ等に用いられる背景からも、魚眼画像に対する水滴除去のニーズは高い。魚眼レンズに水滴が付着した状態で撮影した場合、水滴は大きく歪み、通常のレンズで撮影された画像を対象とする従来の水滴除去手法だけではうまく検出・除去することが困難であると考えられる。

本研究では従来の水滴除去手法である Attentive GAN に加えてテンプレートマッチングにより水滴検出を行っている。魚眼レンズに付着した水滴の形状や水滴に反射した背景を計算して水滴テンプレートを作成し、水滴付着画像とのテンプレートマッチングを行うことにより、魚眼画像から直接水滴検出ができると考えられる。本研究に用いた Deep Fill v2 は Gated Convolution の適用によりどのような形状の画像修復が可能であるため、魚眼レンズに付着した水滴として検出された領域に対して除去が可能である。従って、本研究で提案した手法は魚眼画像を対象とした水滴除去にも有用であると考えられる。

## 謝辞

本研究のテーマや方向性，研究の進め方，細かい知識や技術など多くのご指導いただきました渡辺裕教授と早稲田大学国際情報通信研究センター招聘研究員の原潤一博士に深く感謝申し上げます。

また，日頃から問題解決の相談に乗っていただき，快適な研究環境を作ってくださいました研究室の皆様に御礼申し上げます。

最後に，私が学業に専念できる環境を与え，温かく見守ってくれた家族に心より感謝します。

## 参考文献

- [1]. R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, “Attentive Generative Adversarial Network for Raindrop Removal from a Single Image,” Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.2482-2491, May 2018.
- [2]. Z. Shi, N. Fan, D. Yan, and Q. Chen, “Stereo Waterdrop Removal with Row-wise Dilated Attention,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.3829-3836, Aug. 2021.
- [3]. H. Liao, D. Wang, C. Yang and J. Shin, “Video-based Water Drop Detection and Removal Method for a Moving Vehicle,” Information Technology Journal, 12: pp.569-583, Jan. 2013.
- [4]. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Free-Form Image Inpainting With Gated Convolution,” 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp.4470–4479, Oct. 2019.
- [5]. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.770-778, Jun. 2016.
- [6]. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS), pp.802–810, Dec. 2015
- [7]. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places: a 10 Million Image Database for Scene Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.40, no6, pp.1452-1464, Jun, 2018.
- [8]. 赤松 佳則, “ガラスへの撥水コーティング,” セントラル硝子株式会社 硝子研究所, 2006, Available: <https://www.newglass.jp/mag/TITL/maghtml/82-pdf/+82-p027.pdf>.  
[アクセス日: 15 1 2023]

## 図一覧

図 2.1 Attentive GAN の構造 .....	4
図 2.2 Contextual Autoencoder の構造 .....	5
図 2.3 DeepFill v2 の構造.....	7
図 3.1 Attentive GAN dataset の画像例.....	10
図 3.2 Attentive GAN による水滴除去結果例 .....	11
図 3.3 y 軸方向の輝度変化(1).....	12
図 3.4 y 軸方向の輝度変化(2).....	12
図 3.5 y 軸方向の輝度変化(3).....	13
図 3.6 同マスク領域に対する除去結果の比較.....	14
図 4.1 提案手法の構造.....	16
図 5.1 (左から) 入力画像, Attentive GAN により生成された Attention Map, .....	20
グレースケール化した Attention Map から生成したマスク .....	20
図 5.2 (左から) 入力画像, テンプレートマッチングによる水滴検出結果, .....	21
Attention Map とテンプレートマッチングにより生成されたマスク画像.....	21
図 5.3 (左から) 入力画像, Attentive GAN による水滴除去結果, .....	22
Attentive GAN および DeepFill v2 による水滴除去結果, 提案手法による水滴除去結果, Ground Truth.....	22



## 表一覧

表 3.1 同一のマスク領域に対する除去結果の PSNR および SSIM (5 枚) .....	14
表 5.1 test_a の評価結果 (30 枚) .....	23

## 研究業績

- [1]. M. Arimoto, J. Hara, H. Watanabe, “An Image Inpainting Method Considering Edge Connectivity of Defects”, 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), pp.101-102, October. 2021.
- [2]. T. Shindo, T. Watanabe, R. Yano, M. Arimoto, M. Takahashi, and H. Watanabe, “Super Resolution for QR code Images,” in IEEE Global Conference on Consumer Electronics (GCCE), pp. 281-284, Oct. 2022.
- [3]. 有本麻梨香, 原潤一, 渡辺裕, “GAN および Template Matching を用いた水滴除去手法の検討”, 情報処理学会オーディオビジュアル複合情報処理研究会, Feb. 2023.  
(2023年2月掲載予定)