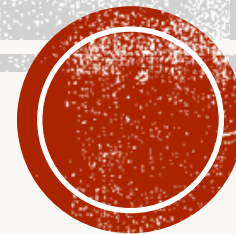
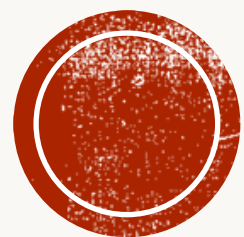


# 生成対抗ネットワークに基づく 多露光画像融合アルゴリズムの 研究と実装

唐朝 (TANG CHAO)



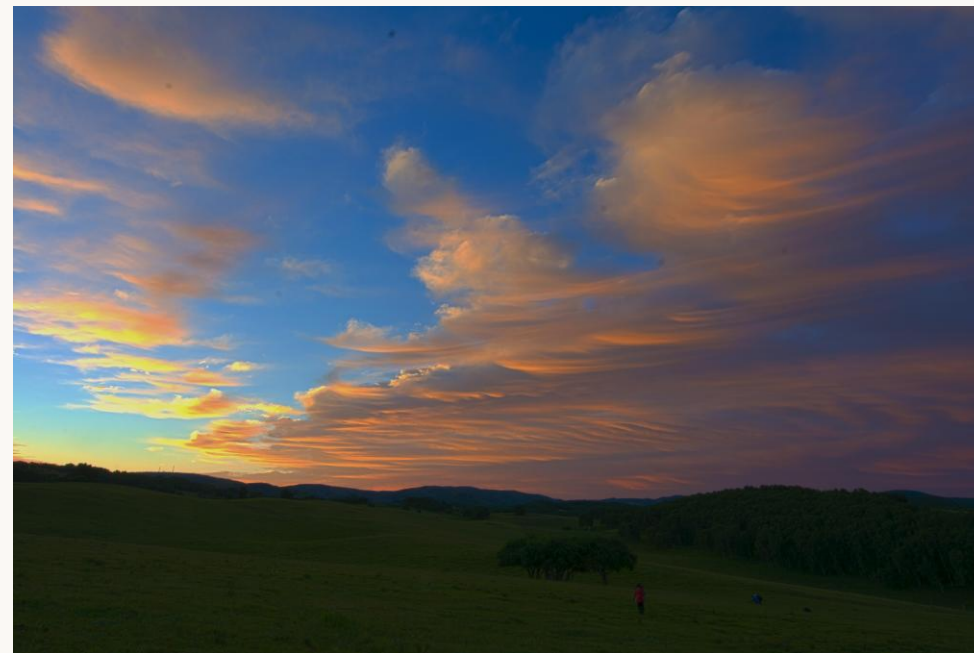


# 研究背景



# 研究背景

日常生活において、私たちは写真撮影を通じて日々の出来事を記録している。スマートフォンの普及に伴い、日常的な撮影の多くはスマートフォンによって行われている。しかし、スマートフォンのハードウェア性能や撮影環境の制約により、画質が十分でない画像が生成される場合がある。



# 研究背景

画質低下の問題を解決するため、撮影時および撮影後の両段階において多くの改善手法が提案されてきた。多露光画像融合アルゴリズムは、露光不足画像および過露光画像の問題を解決するための画像後処理手法の一つであり、ハイダイナミックレンジ（HDR）技術とも呼ばれる。



# 先行研究の進捗

## 従来の多露光画像融合アルゴリズム



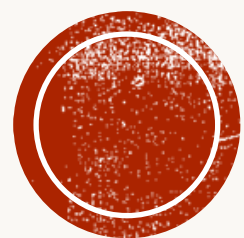
空間領域に基づく手法と変換領域に基づく手法に大別される。空間領域手法は画像の空間的特徴に基づき、空間領域上で直接画像を融合する。一方、変換領域手法では、画像を一度変換領域に分解・表現し、その領域上で融合規則を構築した後、逆変換によって最終的な融合画像を再構成する。



## 深層学習に基づく多露光画像融合アルゴリズム

主に畳み込みニューラルネットワーク（CNN）が用いられており、複数のCNNを用いて生成対抗ネットワーク（GAN）を構築する手法も提案されている。



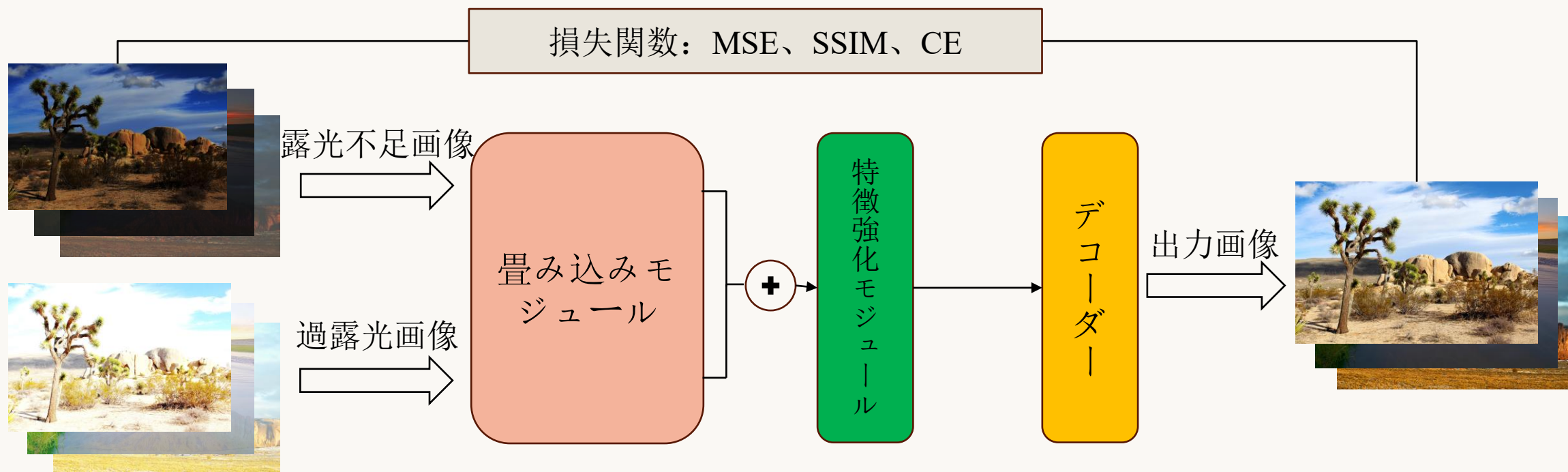


# モデル概要





# モデル概要——生成器



# モデル概要——生成器

畳み込みモジュール



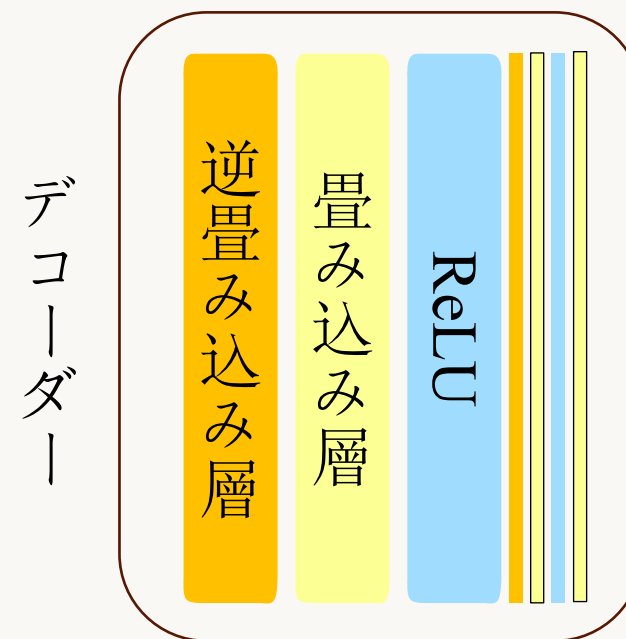
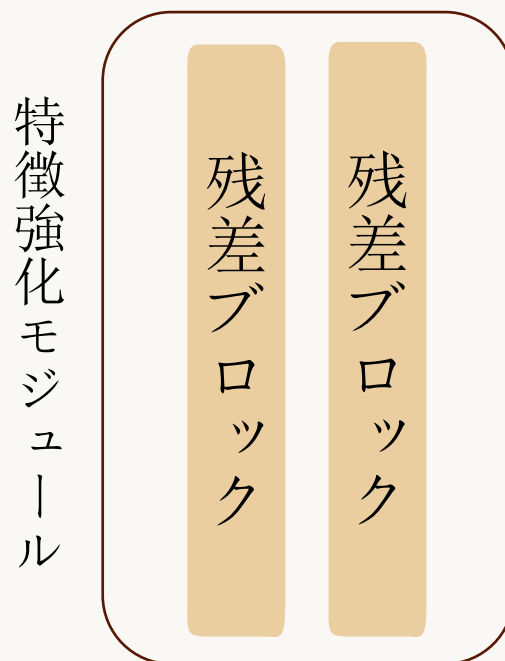
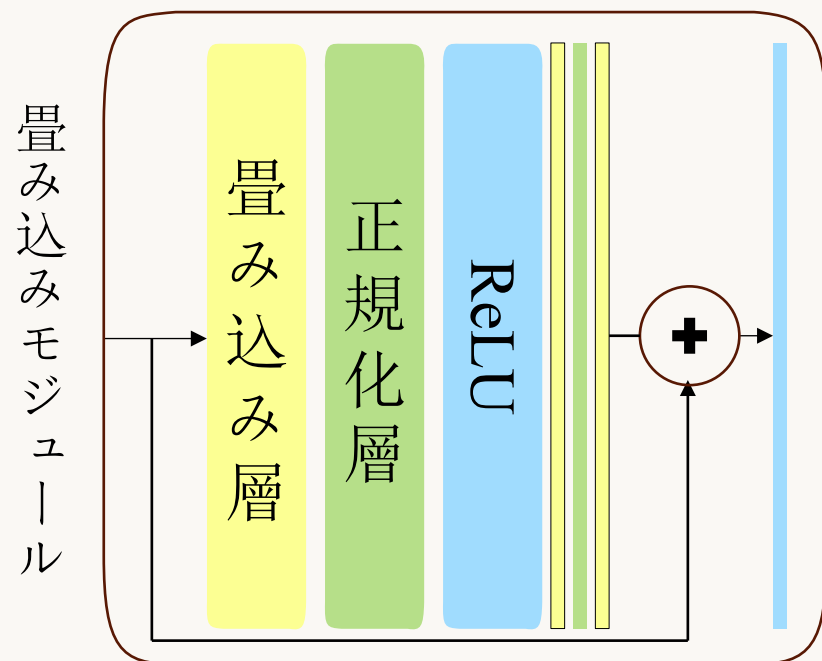
畳み込みモジュールは、残差接続に基づく畳み込みニューラル演算ユニットによって構成されている。残差ネットワークの基本的な考え方は、単一のCNNモジュール内において入力データを出力に直接加算することで、勾配消失問題を緩和すると同時に、ネットワークの深層化による性能向上を可能にする点にある。

残差ネットワークは層数が増加しても性能が低下しにくいため、50層、100層、あるいはそれ以上の深いモデルの構築が可能である。本研究では34層の残差ネットワークを用いて特徴抽出を行い、計算コストを抑えつつ高い性能を実現している。



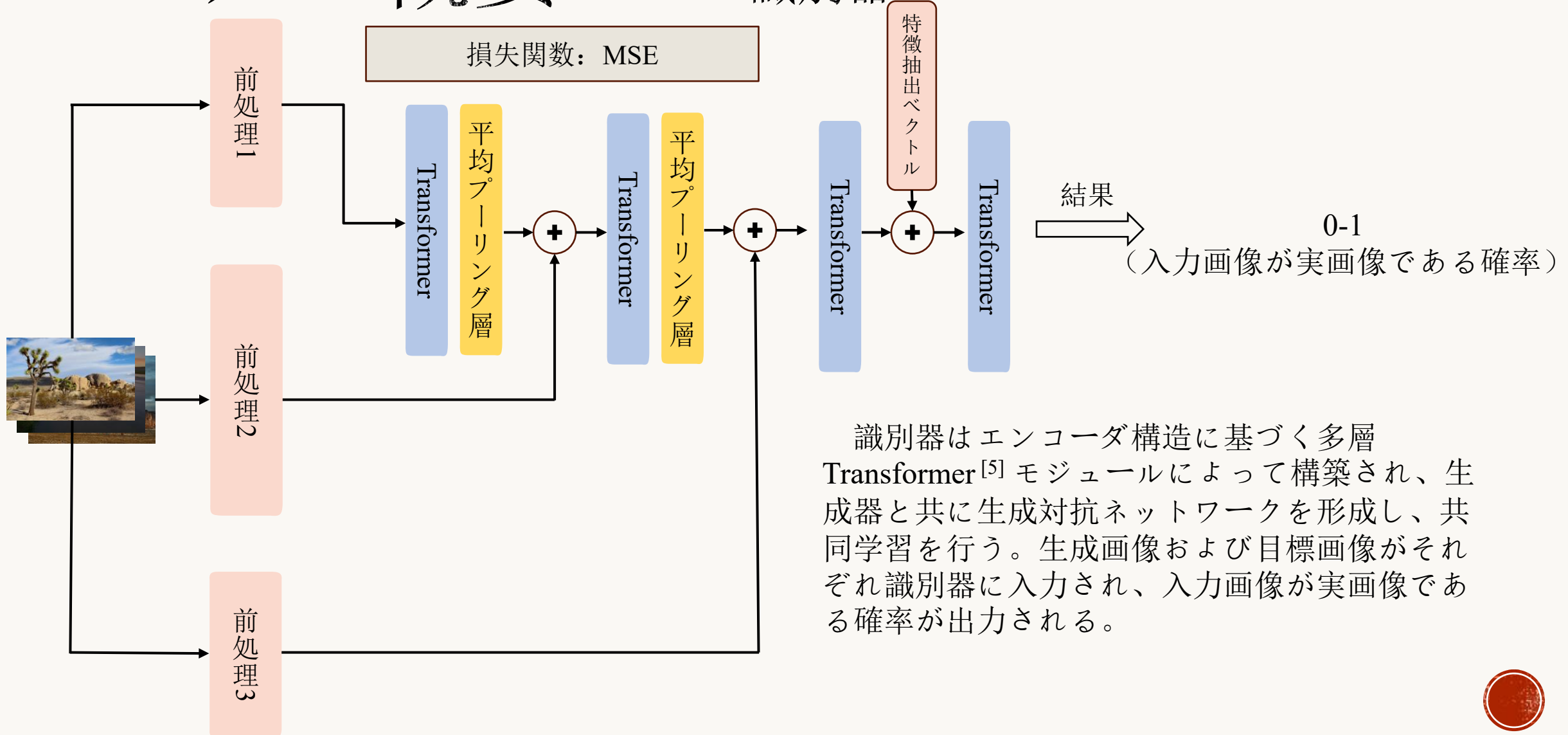


# モデル概要——生成器

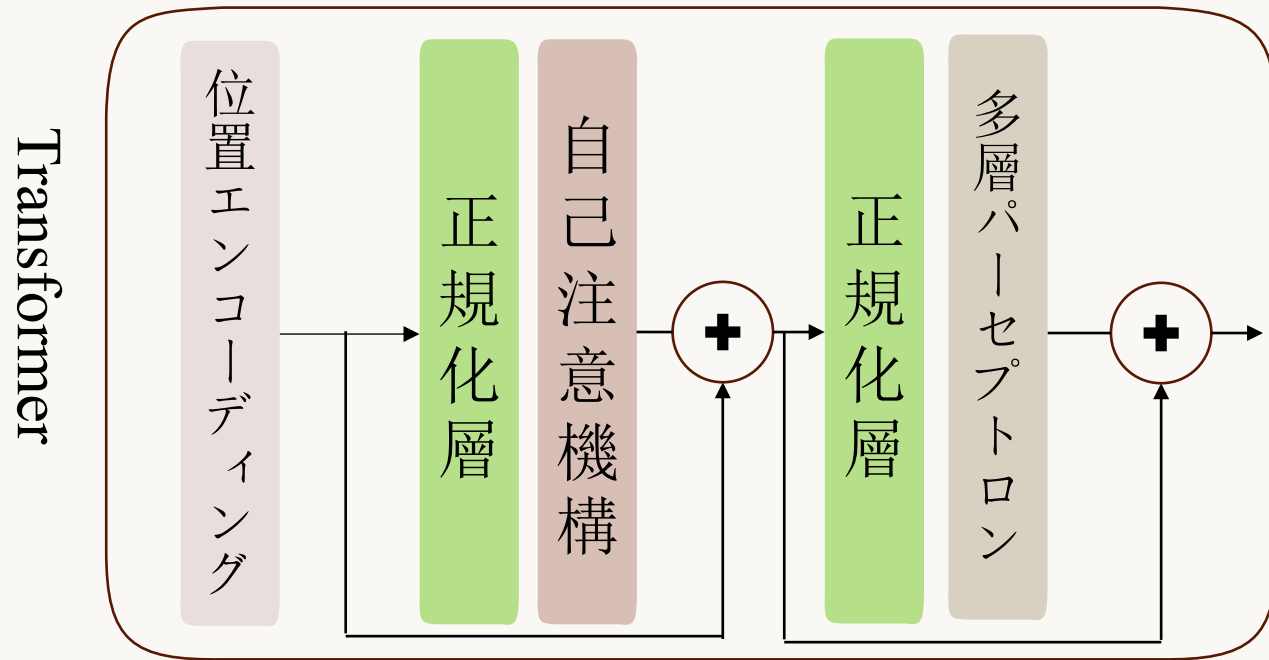


# モデル概要

## 識別器



# モデル概要——識別器



Transformerモジュールは以下の構成を持つ。

位置情報エンコーディングを付加したパッチがまずレイヤ正規化を通過し、その後8ヘッドのマルチヘッド自己注意機構に入力される。得られた出力は元の入力と残差接続され、さらにレイヤ正規化および多層パーセプトロンを経て、再度残差接続を行い最終出力が得られる。



# モデル概要——損失関数

$$L = \alpha L_{SSIM} + \beta L_{MSE} + \gamma L_{CE}$$

$\alpha, \beta, \gamma$  は重み係数である。それぞれは20, 0.1, 5。

$L_{MSE}$  :

$$L_{MSE} = \frac{1}{WHC} \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^C (I_{i,j,k}^{gt} - I_{i,j,k}^F)^2$$

$L_{SSIM}$  :

$$L_{SSIM} = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma$$
$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

$L_{CE}$  :

$$L_{CE} = \frac{1}{N} \sum_i -[y_i \cdot \log_e(p_i) + (1 - y_i) \cdot \log_e(1 - p_i)]$$

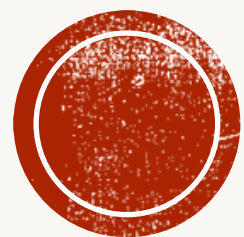


# モデル概要——特徴

本融合アルゴリズムは以下の特性を有する。1) 高精度: 学習済みモデルを用いたリアルタイム処理が可能であり、高度な融合要求に対応できる。2) 高性能: 自己注意機構を用いた識別器により、画像のグローバル特徴を効果的に抽出できる。3) 軽量化: モデル全体の資源消費が少なく、より多くのデバイスへの展開が可能である。

より高い精度および高速な学習を実現するために、本研究では生成器と識別器による対抗学習を導入している。対抗学習では、まず生成器が画像を生成し、その後識別器が生成画像と目標画像を比較する。比較結果は生成器の最適化指標として用いられ、識別器を欺くほど真実味のある画像を生成するよう促される。一方、識別器も同様に比較結果を用いて最適化され、実画像と生成画像をより正確に判別できるようになる。このような反復学習を通じて、生成器と識別器は相互に性能を向上させる。なお、識別器は学習段階のみで使用され、推論段階では生成器のみを用いて融合画像を生成する。





# モデルの実験結果





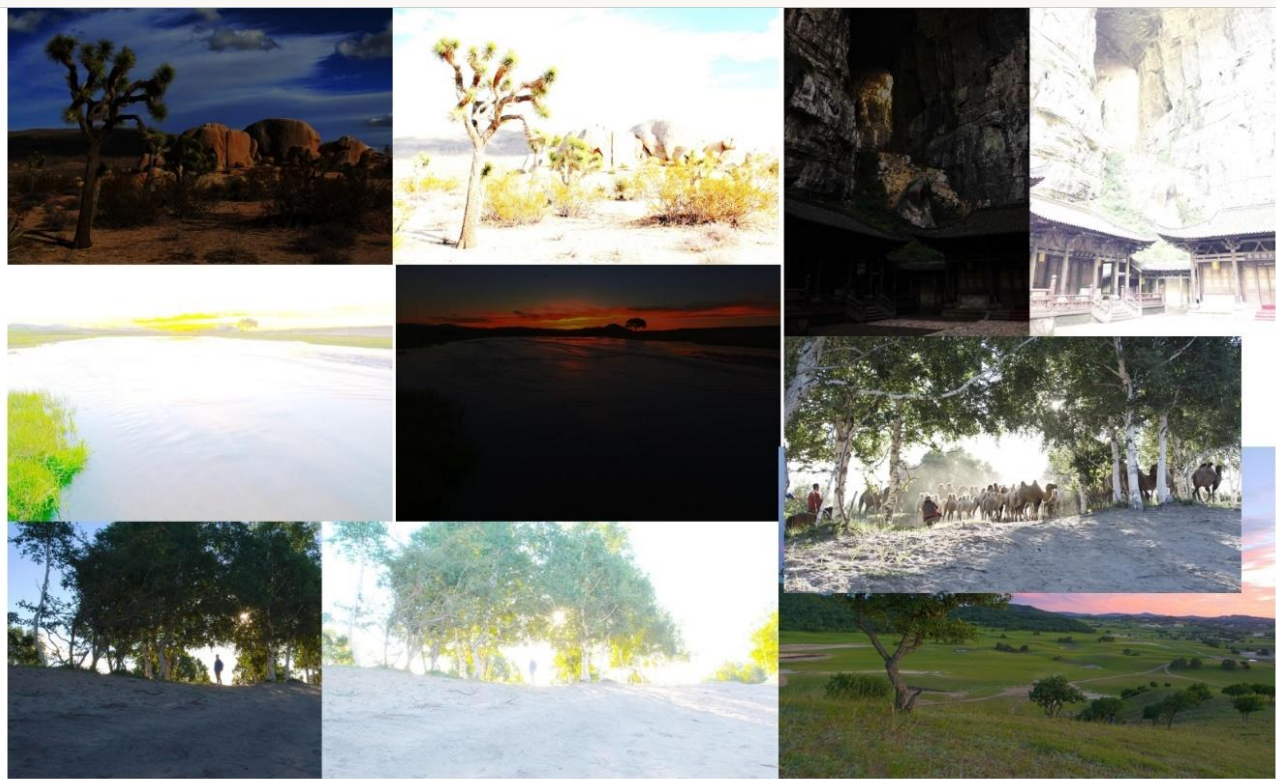
# 学習環境および学習条件

本モデルは Python 言語および TensorFlow 深層学習フレームワークを用いて構築した。ハードウェア環境は Intel® Core™ i7-12700H (2.30GHz) および NVIDIA GeForce RTX 3060 Laptop GPU を搭載し、16GB RAM のノートPCである。OS は Windows 11 (64bit)、実験環境は Python 3.11、TensorFlow 2.12 を使用した。

学習エポック数は1000回とし、生成器には Adam オプティマイザ、識別器には SGD オプティマイザを使用した。初期学習率は 0.001 とし、生成器の学習率は100エポックごとに0.999倍で減衰させ、識別器の学習率は固定とした。バッチサイズは5、総入力画像数は20枚である。



# データセットの概要



本モデルでは、Cai らが提案した多露光画像データセットを用いて学習および評価を行った。

データセットには横向き画像および縦向き画像が混在し、解像度も高いため、計算負荷を軽減する目的で全ての画像を $256 \times 256$ に切り出して処理した。また、入力前にランダムシャッフルを行った。



# 比較手法の概要

本実験では、Mertens ら<sup>[8]</sup>による変換領域に基づく2種類のピラミッド融合手法を比較手法1および2として採用し、Ma ら<sup>[9]</sup>による深層学習に基づく融合手法を比較手法3として選択した。これらと本研究で提案した GAN に基づく融合手法の結果を比較した。まず主観評価を行い、その後複数の評価指標を用いて客観評価を実施した。





# 主観評価

比較結果から、比較手法1は画像がシャープになる一方で歪みが生じ、比較手法2では多くの細部情報が失われていることが確認できる。比較手法3は全体的に暗い画像となっている。

これに対し、本研究で提案した手法はより多くの細部を保持し、原画像に近い結果を得ている。



原画像



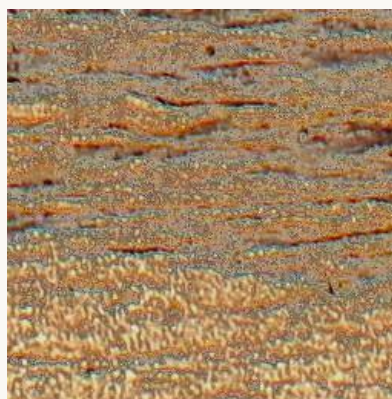
出力画像



比較手法1



入力の露光不足画像



比較手法2



比較手法3



入力の過露光画像





# 主観評価

本グループの結果も前グループと同様の傾向を示しており、比較手法1では歪みが目立ち、比較手法2では細部情報が大きく欠落している。比較手法3は全体的に暗くなる傾向がある。

一方、本研究の手法は細部保持および視覚的品質の両面で優れた結果を示している。



原画像



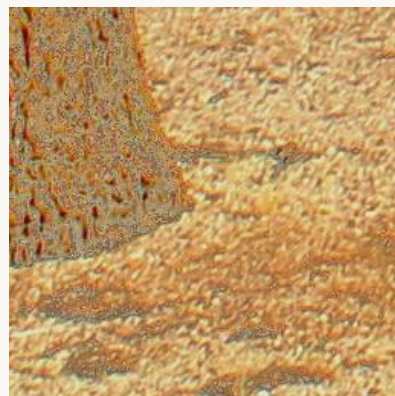
出力画像



比較手法1



入力の露光不足画像



比較手法2



比較手法3



入力の過露光画像



# 客観評価指標の説明

1. 構造類似度（SSIM）：輝度、コントラスト、構造情報の観点から類似度を評価し、1に近いほど高品質である。
2. ピーク信号対雑音比（PSNR）：融合画像の歪みを定量化し、値が大きいほど品質が高い。
3. 平均二乗誤差（MSE）：画素差を評価し、値が小さいほど良好である。
4. 情報エントロピー（SE）：画像が含む情報量を示し、値が大きいほど高品質である。
5. 平均勾配（AG）：テクスチャやエッジ情報の変化を反映し、値が大きいほど良好である。
6. 平均値（MEAN）：画像の平均輝度を示す。
7. 相互情報量（MI）：融合画像が原画像の情報をどれだけ取得しているかを示す指標である。





# 客観評価結果

	SSIM	PSNR	MSE	SE	AG	MEAN	MI
本モデル	0.6813	<b>22.059</b>	<b>482.72</b>	7.284	7.42	<b>498.08</b>	1.2413
比較手法1	<b>0.9253</b>	18.668	1024.84	7.466	12.37	432.13	<b>1.7988</b>
比較手法2	0.6478	15.947	1663.78	7.433	<b>14.83</b>	433.23	1.4897
比較手法3	0.9041	17.175	1246.02	<b>7.658</b>	10.79	407.99	1.1612

評価結果から、本研究の手法は複数の評価指標において最良または準最良の性能を示していることが分かる。高い PSNR と低い MSE を維持しつつ、人間の視覚特性により適合し、高速な融合処理と低い計算資源消費を実現している。



# アブレーション実験

	特徴強化モジュール	識別器
A	√	√
B		√
C	√	
D		

特徴強化モジュールおよび識別器の有無を条件として、以下の3種類のモデルを構築し、元のモデルAと比較した。

モデルB: 特徴強化モジュールを除去

モデルC: 識別器を除去

モデルD: 特徴強化および識別器の両方を除去



# アブレーション実験結果

	SSIM	PSNR	MSE	SE	AG	MEAN	MI
A	<b>0.6813</b>	<b>22.059</b>	<b>482.72</b>	7.284	<b>7.42</b>	<b>498.08</b>	<b>1.2413</b>
B	0.2490	16.728	2308.40	<b>7.479</b>	7.36	464.04	0.8215
C	0.2478	18.984	883.52	7.269	7.20	105.38	0.8490
D	0.6669	19.824	1060.02	7.419	7.19	120.16	1.1071

多くの評価指標において、元のモデルAが最も優れた性能を示した。  
一部の指標では各モデルの結果が近いものの、総合的に判断するとモデルAが最適である。



# ユーザーインターフェース設計



システムの操作性を向上させるため、プログラミング経験のないユーザーでも直感的に操作できる、簡潔なユーザーインターフェースを設計・実装した。



# ユーザーインターフェース設計



操作手順は以下の通りである。

- 1) 露光不足画像および過露光画像を選択する。
- 2) 使用するモデルのパスを選択する。
- 3) 融合処理を開始し、結果を表示する。



# 結論と今後の展望

## 結論



本研究では、静的シーンにおける多露光画像融合問題を対象とし、既存の深層学習手法を改良した。

生成対抗ネットワークに基づく手法が、融合品質および処理効率の両面で有効であることを示した。。

## 今後の展望



現在の融合アルゴリズムには以下の課題が残されている。

- 1) HDR動画への拡張
- 2) 他の画像復元技術との統合
- 3) さらなる計算量削減





# 参考文献

- [1] Yang Z, Chen Y, Le Z, et al. GANFuse: a novel multi-exposure image fusion method based on generative adversarial networks[J]. Neural Computing and Applications, 2021, 33: 6133~6145.
- [2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. Communications of the ACM, 2020, 63(11): 139~144.
- [3] Xu H, Ma J, Zhang X P. MEF-GAN: Multi-exposure image fusion via generative adversarial networks[J]. IEEE Transactions on Image Processing, 2020, 29: 7203~7216.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [6] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [7] Zhou W, Alan C. Bovik, Hamid R. Sheikh, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Trans. Image Process, 2004, 13(4): 600~612.
- [8] Mertens T, Kautz J, van Reeth F. Exposure fusion[C]. 15th Pacific Conference on Computer Graphics and Applications (PG'07), October 29-November 2, 2007, Maui, HI, USA. New York: IEEE Press, 2007: 382- 390.
- [9] Ma K, Duanmu Z, Zhu H, et al. Deep guided learning for fast multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2020, 29: 2808~2819.



ご清聴ありがとうございました。

