AI 技術によるアニメ制作支援：局所物体検出の研究と応用

研究代表者　　鈴木　大三　　筑波大学　システム情報系　准教授

1 はじめに

深層学習の登場以降，画像生成および画像編集技術は著しい進展を遂げている．特に，敵対的生成ネットワーク（GAN）[1] や拡散モデル [2] の登場により，高品質かつ多様な画像生成が可能となり，画像編集やコンテンツ制作の在り方を大きく変革しつつある．これらの技術は，画像全体の生成だけでなく，特定の属性を制御する局所編集にも応用されており，視覚的コンテンツの柔軟な生成を支える基盤技術となっている．

特にアニメ調キャラクター画像においては，目や髪色，表情といった局所的属性を操作することで，キャラクターの印象や意味的情報を大きく変化させることが可能である．このような局所変換は，イラスト制作支援やゲーム・映像制作，さらにはユーザインタラクションの分野において重要な役割を担っており，近年その需要は急速に高まっている．しかしながら，これらの局所変換を高品質かつ安定に実現するためには，変換対象となる領域を正確に特定する必要がある．すなわち，画像中の意味的に適切な領域を抽出する物体検出技術が不可欠である．

物体検出に関しては，これまでに SSD や YOLO 系列に代表される高速かつ高精度な手法が提案されており，実世界画像に対しては高い性能を達成している[3][4]．しかしながら，アニメ調画像は現実画像と比較して形状の抽象化や誇張が顕著であり，テクスチャや色分布も大きく異なるため，従来の物体検出手法をそのまま適用することは困難である．さらに，アニメ画像においては，同一カテゴリ内での形状変動が大きく，また意味的な境界が曖昧であることから，検出対象の定義自体が難しいという問題も存在する．

加えて，本質的な問題として，従来の物体検出手法は，検出精度，例えば mAP の最大化を目的として設計されており，その検出結果が後段の画像生成・変換処理に与える影響についてはほとんど考慮されていない点が挙げられる．実際には，局所画像変換においては，検出領域のわずかなずれや過剰な包含が，生成結果の破綻や不自然さの原因となることが多い．例えば，背景領域の誤検出により不要な変換が適用されたり，対象領域が過剰に広く設定されたりすることで，本来変換すべきでない部分まで影響を受けるといった問題が生じる．このような現象は，従来の評価指標では十分に捉えることができず，検出精度が高いにもかかわらず生成品質が低下するという乖離を引き起こす．

このような背景を踏まえ，本研究では“生成に適した領域検出”という新たな観点を導入する．すなわち，物体検出を単なる認識タスクとしてではなく，画像生成・画像編集のための基盤技術として再定義する．具体的には，画像変換モデルを検出器と統合し，生成結果を評価信号として検出器の学習に直接反映する End-to-End 学習フレームワークを提案する．この枠組みにより，従来の mAP 最適化では得られなかった生成に適した領域抽出を実現することを目的とする．

さらに，本研究の過程において，生成品質を左右する要因として，領域選択だけでなく画像の構造情報，すなわちエッジ，輪郭，周波数成分などが重要であることが明らかとなった．拡散モデルは高品質な画像生成能力を有する一方で，構造制御が難しいという課題が指摘されている[2]．また，低照度画像強調のような画像復元・変換タスクにおいても，局所的な明るさの改善だけでなく，エッジや高周波成分を適切に保持し，不自然なブロック歪みやノイズを抑制することが重要となる．

この課題に対し，本研究では副次的な取り組みとして，二つの構造制御型生成・復元手法を検討した．第一に，エッジ情報を条件として利用する構造制御型画像修復手法を検討した．本手法では，事前修復およびエッジ生成を組み合わせ，洗練されたエッジ情報を拡散モデルに入力することで，構造的一貫性と視覚的自然さの両立を図る．第二に，双直交ウェーブレット変換を利用する周波数成分考慮型低照度画像強調手法を検討した．本手法では，双直交ウェーブレット変換により画像を低周波成分と高周波成分に分解し，各成分の性質に応じた処理を行うことで，処理の高速化を図りつつ，ブロック歪みやノイズを抑制した自然な画像強調を実現する．

以上のように，本研究は，画像生成における“どこを処理するか（領域）”と“どのように構造を保つか（構造）”という二つの観点から問題を捉え，物体検出と画像生成・画像復元の統合的最適化を目指すものである．本報告では，主研究である生成を考慮した物体検出手法に加え，副次的研究として実施した構造制御型画像

公益財団法人電気通信普及財団
研究調査助成報告書　第41号　2026年度

修復手法および周波数成分考慮型低照度画像強調手法を整理し，画像生成基盤技術としての新たな枠組みを提示する．

2 関連研究

2-1 物体検出

物体検出はコンピュータビジョンにおける基本的課題の一つであり，画像中の物体の位置およびカテゴリを同時に推定する問題として広く研究されてきた．従来は，特徴量設計に基づく手法が主流であったが，近年では深層学習の発展に伴い，畳み込みニューラルネットワーク（CNN）を用いた手法が主流となっている．

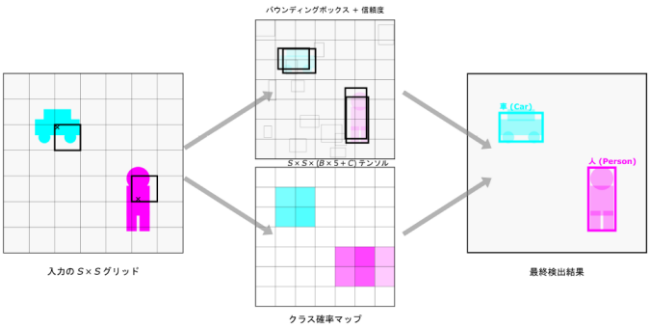

図 1．YOLO における検出プロセス．

代表的な手法として，領域提案に基づく二段検出器（Faster R-CNN など）[5] と，単段検出器（SSD [3], YOLO 系列 [4] など,図 1 参照）が挙げられる．二段検出器は高精度な検出を実現する一方で計算コストが高く，リアルタイム処理には不向きである．一方，単段検出器は領域提案を省略することで高速な推論を可能とし，実用的な応用において広く利用されている．特に YOLO 系列は，検出精度と速度のバランスに優れ，多様な応用分野に展開されている．

しかしながら，これらの手法はいずれも検出精度の最大化を目的として設計されており，評価指標としては主に mAP（mean Average Precision）が用いられている．この評価基準は，検出ボックスと正解ボックスの重なり（IoU）に基づいており，認識タスクとしては合理的であるものの，検出結果の利用価値を直接反映するものではない．特に，後段に画像生成や画像編集を伴う場合，検出領域のわずかなずれや過剰な包含が生成結果に大きな影響を及ぼす可能性があるにもかかわらず，そのような観点は従来の評価枠組みには含まれていない．

さらに，多くの検出手法では非極大抑制（Non-Maximum Suppression; NMS）が用いられている．NMS は冗長な検出結果を除去するための重要な処理であるが，非連続な操作であるため勾配計算が困難であり，他のニューラルネットワークモジュールとの End-to-End 学習を阻害する要因となる．この問題に対して，近年では NMS を不要とする設計や，より連続的な最適化を可能とする手法も提案されているが[6]，生成タスクとの統合という観点からの検討は依然として限定的である．

以上のように，従来の物体検出研究は主として認識精度の向上に焦点を当てており，生成処理との整合性を考慮した設計は十分に行われていない．本研究では，この点に着目し，生成結果を考慮した新しい物体検出の枠組みを提案する．

2-2 画像生成・画像変換

画像生成および画像変換の分野では，深層生成モデルの発展により飛躍的な進展が見られている．特に，Generative Adversarial Networks（GAN）[1] は，高品質な画像生成を可能とする代表的手法であり，画像変換やスタイル変換に広く応用されている．CycleGAN [7] はその代表例であり，対応関係のない画像間の変換を可能とすることで，多様な画像編集タスクに応用されている（図 2 参照）．

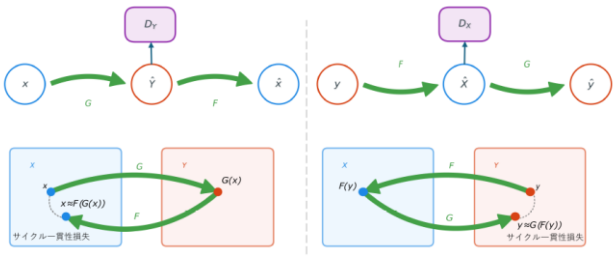

図 2．CycleGAN の概要図．

一方，近年では拡散モデル [2] が注目を集めており，逐次的なノイズ除去過程を通じて高品質かつ安定した画像生成を実現している（図 3 参照）．拡散モデルは, GAN と比較して学習の安定性やモード崩壊の回避といった利点を有しており，画像生成のみならず画像修復や超解像などのタスクにも応用が進

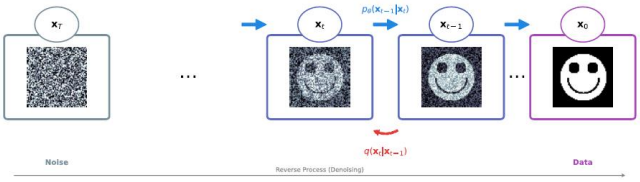

図 3．拡散モデルの概念図．

んでいる．

しかしながら，これらの生成モデルにはそれぞれ課題が存在する．GAN は高い表現能力を有する一方で学習が不安定であり，アーティファクトが発生しやすい．一方，拡散モデルは高品質な生成が可能であるが，入力条件に対する構造的制約を明示的に扱うことが難しく，大域的な整合性が崩れる場合がある．

画像修復においても同様の課題が存在する．従来の手法としては，エッジ情報を利用して構造を補完する手法（EdgeConnect [8]）や，文脈情報に基づく補完を行う手法が提案されているが，構造的一貫性と視覚的自然さの両立は依然として難しい問題である．このため，構造情報を適切に制御しながら生成を行う手法の開発が重要な課題となっている．

2-3 検出と生成の統合

物体検出と画像生成は従来独立した研究分野として扱われてきたが，実際の応用において，両者は密接に関連している（図 4 参照）．特に，局所編集を伴う画像生成では，検出された領域に対して変換を適用するため，検出精度が生成品質に直接影響を及ぼす．しかしながら，従来の多くの研究では，物体検出と画像生成は別々に最適化されており，両者の相互作用は考慮されていない．この結果，検出精度が高いにもかかわらず生成結果が不自然になるといった問題が生じる．このような問題に対し，複数タスクを統合する End-to-End 学習やマルチタスク学習の研究が進められているが，生成結果を直接評価信号として物体検出器の学習に組み込む枠組みはほとんど検討されていない．

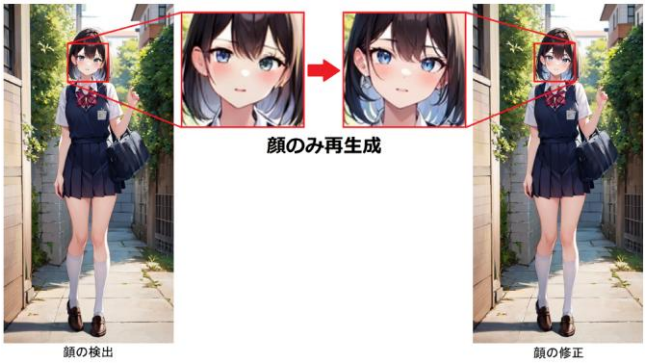

図 4．局所物体検出を用いた局所画像変換の

本研究では，この点に着目し，画像変換結果を評価信号として利用することで，生成に適した領域検出を実現する新たなフレームワークを提案する．これにより，従来の検出精度中心の枠組みを超えた，新しい最適化基準に基づく物体検出を実現する．

3 提案手法（主研究）

3-1 フレームワーク概要

本研究では，物体検出と画像生成を統合した End-to-End 学習フレームワークを提案する．従来の物体検出は認識精度の最大化を目的として設計されているのに対し，本研究では生成に適した領域抽出を目的とする．そのために，画像変換結果を評価信号として検出器の学習に直接反映する枠組みを導入する．

入力画像を $\mathbf{I}$，物体検出器を$D(\cdot)$，画像変換モデルを $G(\cdot)$ とする．検出器により得られる領域マスクを $\mathbf{M} = D(\mathbf{I})$ とすると，変換後画像$\hat{\mathbf{I}}$ は以下のように定義される：

$$\hat{\mathbf{I}} = \mathbf{M} \odot G(\mathbf{I}) + (\mathbf{1} - \mathbf{M}) \odot \mathbf{I}$$

ここで，$\odot$ は画素ごとの積を表す．すなわち，検出領域に対してのみ変換を適用し，それ以外の領域は元画像を保持する．

本手法の処理の流れは以下の通りである：

1. 入力画像 $\mathbf{I}$ に対して物体検出器 D を適用し，対象領域マスク $\mathbf{M}$ を取得する
2. 画像変換モデル G を用いて変換画像 $G(\mathbf{I})$ を生成する
3. マスク $\mathbf{M}$ に基づいて局所変換画像 $\hat{\mathbf{I}}$ を構成する
4. $\hat{\mathbf{I}}$ を教師データまたは自己教師信号と比較し，損失を計算する
5. 損失を検出器 D に逆伝播することで，生成に適した領域抽出を学習する

このように，生成結果を介して検出器を最適化することで，従来の mAP 最適化とは異なる観点からの領域抽出が可能となる．

3-2 モデル構成

提案手法では，物体検出器と画像変換モデルを統合した構成を採用する．

物体検出器には，非極大抑制（NMS）を必要としない YOLOv10 [6] を用いる．YOLOv10 は，NMS を排除した End-to-End な検出を実現しており，検出過程全体が微分可能である．これにより，検出結果に基づく生成誤差を検出器に対して直接逆伝播することが可能となる．従来の NMS ベースの検出器では，このような統合は困難であった．

画像変換モデルには CycleGAN [7] を採用する．CycleGAN は対応関係のない画像間の変換を可能とし，スタイル変換や属性変換において高い性能を示す．本研究では，目や髪色などの局所的属性変換を対象とし，検出された領域に対してのみ変換を適用することで，局所編集を実現する．

このように，検出器と生成モデルを統合することで，“どこを変換するか”と“どのように変換するか”を同時に最適化する枠組みを構築する．

3-3 損失関数

本研究では，生成品質を考慮した検出を実現するために，以下の二つの損失関数を導入する．

(1) 領域外安定性損失

検出対象外の領域において不要な変化が生じないよう制約する損失である（図 5 参照）．誤検出により背景領域が変換されることを抑制する役割を持ち，以下のように定義される：

$$L_{\mathrm{LROS}} = ||(1 - \mathbf{M}) \odot (G(\mathbf{I}) - \mathbf{I})||_1$$

この損失は，非検出領域における変換量を最小化することで，検出領域の過剰な拡大を抑制する．すなわち，背景領域を含むような誤検出に対して強いペナルティを与える．

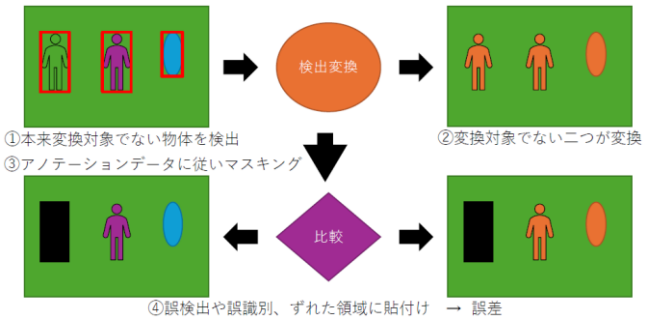

図 5. 領域外安定性損失の概念図.

(2) 領域一貫性損失

検出された領域に対して適切な変換が行われることを促す損失である（図 6 参照）．対象領域が適切に抽出されていない場合，変換結果に歪みや不自然さが生じるため，本損失により検出領域の精度が改善される．教師信号となる理想的な変換画像を $\mathbf{I}^*$ とすると，以下のように定義される：

$$L_{\mathrm{LRC}} = ||\mathbf{M} \odot (\hat{\mathbf{I}} - \mathbf{I}^*)||_1$$

この損失は，検出領域内における変換結果の誤差を評価するものであり，対象領域の位置および形状が適切であるほど損失が小さくなる．

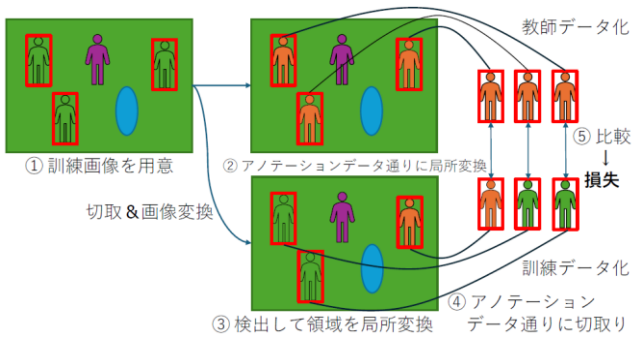

図 6. 領域一貫性損失の概念図.

(3) 全体損失

最終的な損失関数は，従来の検出損失 L_{det} と提案損失を統合した以下の形で定義される：

$$L = L_{\mathrm{det}} + \lambda_1 L_{\mathrm{LROS}} + \lambda_2 L_{\mathrm{LRC}}$$

ここで，λ_1, λ_2 は重み係数である．

3-4 学習の特徴

提案手法の最大の特徴は，物体検出を認識精度ではなく生成品質に基づいて最適化する点にある．従来の検出手法では，IoU に基づく位置精度のみが評価対象であったが，本手法では色やテクスチャといった生成特性が損失関数を通じて反映される．

これにより，検出領域は単に正解ボックスに一致するだけでなく，生成処理にとって最適な領域へと適応的に変化する．具体的には，背景領域の包含を避けるために検出領域が縮小する傾向や，変換対象の境界に沿った精密な領域抽出が促進されるといった挙動が観察される．

また，YOLOv10 を用いることで検出から生成までの完全な End-to-End 学習が可能となり，従来のパイ

プライン型処理では実現できなかった統合最適化を達成している．

以上により，本提案手法は，画像生成と物体検出の関係を再定義し，生成品質を考慮した新たな物体検出の枠組みを提供する．

4 実験

4-1 データセット

本研究では，アニメ調キャラクター画像を対象とした独自データセットを構築した（図 7 参照）．本データセットは，画像生成ツールおよび既存の公開データを組み合わせて収集した約 17,000 枚以上の画像から構成される．各画像に対して，目や髪色などの局所的属性に関するアノテーションを付与し，物体検出および局所変換タスクの双方に利用可能な形式とした．アニメ調画像は，現実画像と比較して形状のデフォルメや色分布のばらつきが大きく，また意味的境界が曖昧であるため，アノテーションの付与には慎重な設計が必要となる．本研究では，属性ごとに検出対象領域の定義を明確化し，一貫した基準に基づいてラベル付けを行った．また，データの多様性を確保するため，キャラクターの姿勢，視点，表情，照明条件などが偏らないように収集を行っている．データセットは学習用，検証用，評価用に分割し，モデルの過学習を防ぐとともに汎化性能の評価を可能とした．

図 7．データセット画像の一例．

4-2 評価指標

評価指標としては，従来の物体検出性能を測る指標である mAP@0.5 を用いた．これは，検出ボックスと正解ボックスの IoU（Intersection over Union）が 0.5 以上である場合に正解とみなす指標であり，物体検出の標準的な評価方法である．しかしながら，本研究の目的は単なる検出精度の向上ではなく，生成に適した領域抽出の実現であるため，mAP のみでは評価が不十分である．そのため，本研究では生成結果の品質についても定性的評価を行った．具体的には，検出領域に基づいて適用された画像変換の結果を観察し，対象領域に対する変換の適切性，背景領域における不要な変換の有無，および境界付近における不自然さやアーティファクトの発生の観点から評価を行った．これにより，従来の評価指標では捉えられない生成品質への影響を総合的に評価した．

4-3 実験設定

提案手法の有効性を検証するため，以下の 2 種類の手法との比較を行った：

1. ベースライン検出器（従来の検出損失のみで学習）
2. 提案手法（生成損失を統合した End-to-End 学習）

両手法において，同一のネットワーク構造および学習データを用い，損失関数のみを変更することで，提案損失の効果を明確に評価した．学習には一般的な最適化手法（Adam 等）を用い，学習率やバッチサイズなどのハイパーパラメータは検証データに基づいて調整した．また，提案損失における重み係数 λ_1, λ_2 は，生成品質と検出精度のバランスを考慮して設定した．

4-4 実験結果

実験の結果，提案手法はベースライン手法と比較して，mAP@0.5 において一定の向上が確認された（表 1 参照）．これは，生成結果を評価信号として利用することで，検出器がより適切な領域を学習したことを示し

ている．

さらに重要な点として，提案手法では検出領域が従来よりもコンパクトになる傾向が観察された．従来の物体検出では，対象物体全体を包含するように検出ボックスが設定されることが望ましいとされてきたが，本研究の結果は，生成タスクにおいては必ずしもその限りではないことを示唆している．すなわち，生成に適した領域は，変換対象に必要な最小限の範囲に限定される方が望ましく，過剰な領域の包含はむしろ生成品質の低下を招く要因となる．

表 1．各眼の色の検出精度比較．

Target Color	# Images	Max mAP@0.5	Final mAP@0.5	Precision	Recall
Green	200	0.880	0.874	0.813	0.799
Blue	**200**	**0.732**	**0.717**	**0.698**	**0.634**
Purple	200	0.717	0.717	0.672	0.701
Red	200	0.618	0.581	0.601	0.597
Yellow	200	0.602	0.583	0.592	0.628
Pink	200	0.589	0.578	0.560	0.555
Orange	200	0.465	0.422	0.440	0.588
Black	137	0.039	0.004	0.009	0.071
Grey	112	0.000	0.000	0.000	0.250

また，定性的評価においても，提案手法は，背景領域における不要な変換の抑制，対象領域の境界に沿った自然な変換，および変換結果の視覚的一貫性の向上の点で優れた結果を示した．これらは，領域外安定性損失および領域一貫性損失が有効に機能していることを示している．

さらに，学習過程においても，提案手法は初期段階から安定した収束挙動を示した．従来手法では，誤検出に起因する不安定な更新が見られる場合があるのに対し，本手法では生成結果を介したフィードバックにより，検出器の更新がより一貫した方向に導かれることが確認された．

以上の結果から，提案手法は従来の検出精度指標を超えた新たな観点において有効であり，生成タスクに適した物体検出を実現する有望なアプローチであるといえる．

5 副次的研究：構造制御型生成・復元手法

5-1 構造制御型画像修復

本研究の過程において，画像生成における品質を左右する要因として，“どの領域を処理するか”に加えて，“どのような構造を維持・再構成するか”が極めて重要であることが明らかとなった．特に，拡散モデルを用いた画像生成においては，高い生成能力を有する一方で，輪郭やエッジといった構造情報の明示的な制御が難しく，欠損領域が大きい場合には形状の崩れや不自然なテクスチャが生じる場合がある．

この課題に対し，本研究では副次的な取り組みとして，エッジ情報を条件として利用する構造制御型画像修復手法を検討した（図 8 参照）．本手法は，事前修復，エッジ洗練，拡散モデルによる画像修復の三段階から構成される．事前修復により画像全体の大まかな構造および色分布を補完し，その結果からエッジ情報を抽出・洗練することで，後段の拡散モデルに対して構造的な条件情報を与える．これにより，拡散モデルの高い生成能力を活かしつつ，欠損領域における輪郭や線構造の一貫性を維持した画像修復を実現することを目的とした．

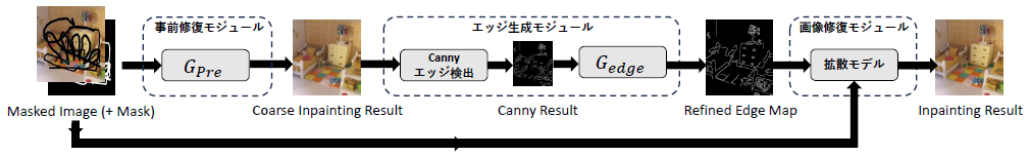

図 8．提案した画像修復手法のフロー．

第一段階の事前修復モジュールでは，入力画像と欠損マスクを用いて粗い修復画像を生成する．この段階では，欠損領域を完全に高精度に復元することよりも，後段のエッジ抽出に利用可能な大域的構造と色分布を得ることを目的とする．第二段階のエッジ洗練モジュールでは，事前修復画像に基づいて得られた粗いエッジ情報を入力とし，学習ベースのエッジ生成モデルにより，欠損領域を含む連続性の高いエッジマップを生成する．第三段階では，欠損画像，マスク，および洗練されたエッジ情報を条件として拡散モデルに入力し，逐次的なノイズ除去過程を通じて最終的な修復画像を生成する．

提案手法の有効性を検証するため，Places2 データセットを用いた比較実験を行った（表 2 および図 9 参照）．比較対象として，EdgeConnect [8]，AOT-GAN [9]，RePaint [10]を用い，欠損率 20%，30%，40%，

表 2. 画像修復結果の定量評価.

評価指標	欠損率 [%]	EC	AOT	Repaint	Ours
PSNR [dB] ↑	20	29.00	27.38	**30.74**	30.29
	30	26.22	25.69	**27.90**	26.74
	40	24.06	23.39	**25.10**	23.46
	50	21.94	21.63	**22.50**	19.90
SSIM ↑	20	0.932	0.913	0.933	**0.955**
	30	0.883	0.873	0.895	**0.920**
	40	0.824	0.816	0.840	**0.877**
	50	0.748	0.750	0.743	**0.821**
FID ↓	20	12.92	23.10	17.01	**11.82**
	30	20.91	26.56	28.24	**20.43**
	40	**31.42**	35.85	42.71	31.49
	50	**45.16**	47.08	60.22	65.18

図 9. 画像修復結果の定性評価（最右列が提案法）.

50%の条件で評価した．その結果，提案手法は SSIM において全欠損条件で最良の性能を示し，構造的一貫性の向上が確認された．また，FID についても 20%および 30%の中程度の欠損条件において最良の結果を示した．定性的にも，提案手法は輪郭や線構造を滑らかに接続し，中程度までの欠損条件において自然かつ一貫した修復結果を実現した．

一方で，欠損率が 40%および 50%と大きい条件では，PSNR や FID において一部の従来手法が優位となる場合も確認された．このことから，提案手法は特に中程度までの欠損領域における構造保持に有効である一方，大規模欠損に対する生成安定性には改善の余地があると考えられる．以上より，エッジ情報を拡散モデルの条件として利用することは，画像修復における構造的一貫性と視覚的自然さの両立に有効であることが示された．

5-2 周波数成分考慮型低照度画像強調

さらに，本研究では，構造情報を考慮した画像復元・変換の一例として，双直交ウェーブレット変換を利用する低照度画像強調手法についても検討した．低照度画像強調では，画像全体の明るさやコントラストを改善するだけでなく，エッジやテクスチャなどの高周波成分を保持しつつ，ノイズやブロック歪みを抑制することが重要である．特に，拡散モデルを用いた従来手法では，高品質な明るさ補正が可能である一方，ウェーブレット変換の設計や高周波成分の復元が不十分な場合，ブロック状のアーティファクトや過平滑化が生じるという課題がある．

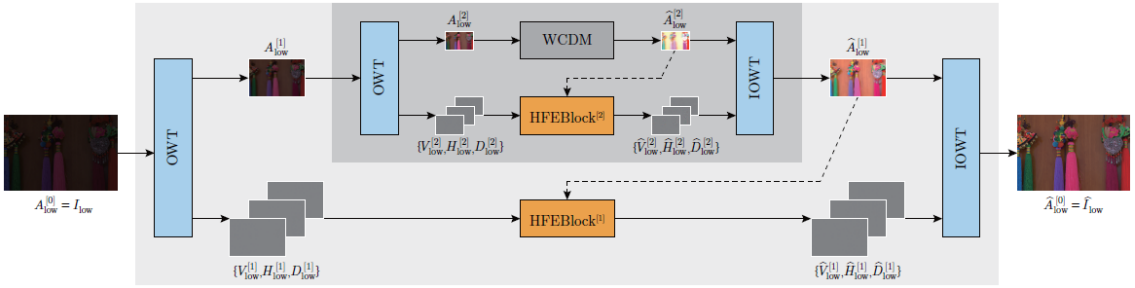

図 10. 提案した低照度画像強調手法のフロー.

この課題に対し，本研究では Overlapped Wavelet Diffusion（OWDiff）を提案した（図 10 参照）．OWDiff は，既存の拡散モデルベースの低照度画像強調手法である DiffLL [11]を基盤とし，その構造的な問題を改善するものである．DiffLL では Haar ウェーブレット変換を用いて画像を低周波成分と高周波成分に分解し，低周波成分に対して拡散モデルを適用することで計算効率を高めている．しかし，Haar ウェーブレット変換は 2×2 ブロック単位の非重複変換であるため，ブロック境界に不連続性が生じやすく，復元画像にブロック歪みが現れる場合がある．また，高周波復元モジュールの能力が十分でない場合，エッジや細部構造が過度に平滑化される．

OWDiff では，この問題を解決するために，Haar ウェーブレット変換の代わりに双直交ウェーブレット変換に基づく重複型ウェーブレット変換を導入した．具体的には，Bior2.2 および Bior4.4 に基づく Overlapped Wavelet Transform（OWT）を用いることで，隣接領域間の相関を考慮した周波数分解を行い，ブロック境界における構造的不連続を抑制する．これにより，非重複変換に起因するブロック歪みを構造的に低減することが可能となる．

一方で，重複型ウェーブレット変換では，高周波成分の一部が低周波成分側に再配分されるため，細部構造が弱まる可能性がある．そこで本研究では，低周波成分をガイドとして高周波成分を補正する High-Frequency Enhance Block（HFEBlock）を導入した．HFEBlock は，拡散モデルにより強調された低周波成分を手掛かりとして，高周波成分に含まれるエッジやテクスチャを選択的に復元する．これにより，OWT によるブロック歪み抑制と，HFEBlock による細部構造復元を相補的に組み合わせ，構造的一貫性と細部保持を両立する低照度画像強調を実現した．

表 3．低照度画像強調結果の定量評価．

Method	LOLv1				LOLv2-real			
	PSNR [dB] ↑	SSIM ↑	LPIPS ↓	Time [spi] ↓	PSNR [dB] ↑	SSIM ↑	LPIPS ↓	Time [spi] ↓
SNR-Aware [20]	<u>23.19</u>	0.800	0.344	<u>0.026</u>	24.17	0.832	0.286	<u>0.030</u>
Retinexformer [21]	21.81	0.808	0.360	**0.023**	26.92	0.860	0.212	**0.025**
GSAD [22]	22.32	**0.848**	0.357	0.425	24.77	0.891	0.260	0.438
DiffLL (K = 1) [23]	22.75	0.831	0.344	0.101	28.01	0.887	0.185	0.103
DiffUIR [24]	21.96	0.824	0.350	0.234	24.67	0.871	0.260	0.241
Wave-Mamba [25]	22.61	0.827	0.332	0.038	25.07	0.872	0.237	0.042
OWDiff (Bior2.2, K = 1)	23.11	0.844	<u>0.322</u>	0.104	<u>28.28</u>	<u>0.897</u>	**0.172**	0.105
OWDiff (Bior4.4, K = 1)	**23.50**	<u>0.847</u>	**0.320**	0.106	**28.41**	**0.899**	<u>0.176</u>	0.108

*The first and second rankings are bolded and underlined, respectively.

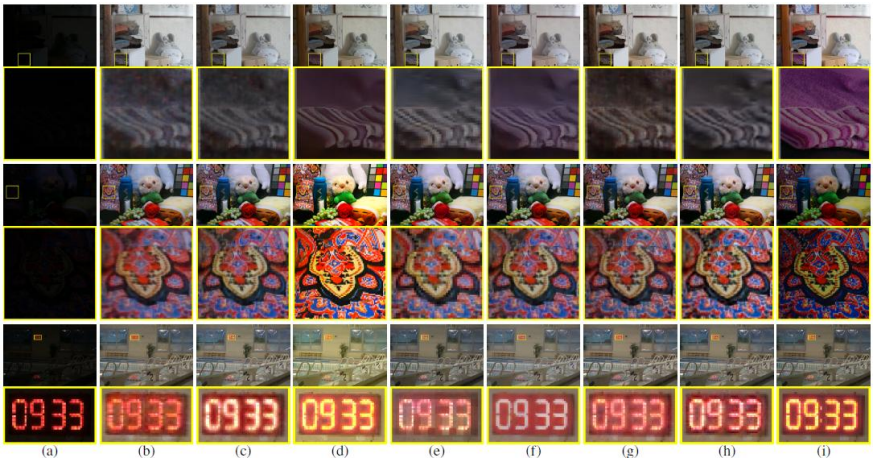

図 11．低照度画像強調結果の定性評価（最右列が提案法）．

実験では，LOLv1 および LOLv2-real データセットを用いて，DiffLL の他，SNR-Aware [12]，Retinexformer [13]，GSAD [14]，DiffUIR [15]，Wave-Mamba [16] などの既存手法と比較した（表 3 および図 11 参照）．その結果，OWDiff は DiffLL と比較して，PSNR，SSIM，LPIPS の各指標において一貫した改善を示した．特に，Bior4.4 を用いた OWDiff では，LOLv1 において PSNR 23.50 dB，SSIM 0.847，LPIPS 0.320 を達成し，DiffLL の PSNR 22.75 dB，SSIM 0.831，LPIPS 0.344 を上回った．また，LOLv2-real においても，PSNR 28.41 dB，SSIM 0.899，LPIPS 0.176 を示し，DiffLL よりも高い復元性能を確認した．

さらに，定性的評価においても，OWDiff は DiffLL で見られたブロック歪みを抑制し，過度な平滑化を避けながらエッジやテクスチャを自然に復元できることが確認された．アブレーション実験からは，OWT がブロック歪みの抑制に有効であり，HFEBlock が高周波成分の復元に寄与することが示された．これらの結果から，低照度画像強調においても，空間領域だけでなく周波数領域における構造情報を適切に扱うことが，生成品質および復元品質の向上に重要であることが明らかとなった．

5-3 主研究との関係

本章で述べた二つの副次的研究は，主研究である生成に適した領域検出と密接に関連している．主研究では，局所画像変換において“どの領域を処理するか”という空間的選択が重要であることに着目し，物体検出を画像生成・画像編集のための基盤技術として再定義した．これに対し，構造制御型画像修復および周波数成分考慮型低照度画像強調では，“処理対象領域の内部でどのような構造を維持・復元するか”という観点に焦点を当てている．

構造制御型画像修復では，エッジ情報を条件として利用することで，欠損領域における輪郭や線構造の一

貫性を保つことを目指した．一方，周波数成分考慮型低照度画像強調では，画像を低周波成分と高周波成分に分解し，明るさや大域的構造を担う低周波成分と，エッジやテクスチャを担う高周波成分をそれぞれ適切に処理することで，ブロック歪みの抑制と細部構造の保持を実現した．これらはいずれも，画像生成・画像復元において構造情報を明示的に扱うことの重要性を示すものである．

以上より，本研究全体は，領域と構造という二つの観点から画像生成・画像編集・画像復元を統合的に捉えるものである．すなわち，主研究では生成に適した領域を検出することで“どこを処理するか”を扱い，副次的研究ではエッジ情報および周波数成分を利用することで“どのように構造を保つか”を扱った．これらの成果は，アニメ調画像の局所編集に限らず，画像修復，低照度画像強調，および各種画像生成タスクにおいて，生成品質を向上させるための基盤的知見を提供するものである．

6 考察・今後の課題

本研究の最も重要な貢献は，物体検出の目的を従来の“認識精度の最大化”から，“生成処理に適した領域抽出”へと再定義した点にある．この視点の転換により，物体検出を画像生成の前処理ではなく，生成品質を左右する基盤技術として位置付ける新たな枠組みを提示した．

提案手法により，生成処理に適した領域抽出に関するいくつかの知見が得られた．まず，局所画像変換においては，検出領域は必ずしも大きいほど良いわけではなく，生成に必要な最小限の範囲に限定することが望ましいことが確認された．過剰に広い領域を検出すると，背景領域まで変換対象に含まれ，不要な色変化や境界付近のアーティファクトが生じるため，結果として生成品質の低下につながる．一方，生成結果を評価信号として検出器の学習に反映することで，従来の mAP 最適化では捉えられなかった，生成品質を考慮した新たな最適化基準を導入できることが示された．これらの結果は，物体検出と画像生成の関係を再考する上で重要な示唆を与えるものであり，特に局所編集を伴う生成タスクにおいて有効であると考えられる．

一方で，本研究にはいくつかの課題も残されている．まず，提案手法は検出と生成を統合した End-to-End 学習を行うため，計算コストが高く，大規模データセットへの適用が容易ではない．特に，拡散モデルなど高計算量の生成モデルを組み合わせた場合，学習時間およびメモリ使用量の増大が課題となる．

また，アニメ調画像におけるアノテーションは主観的要素を含むため，ラベルの曖昧性が学習結果に影響を与える可能性がある．さらに，データセットの分布や属性の偏りが，モデルの汎化性能に影響を及ぼす点についても検討が必要である．

加えて，本手法は画像変換モデルの性能に依存するため，生成モデルの選択および設計が重要な要素となる．より高精度な生成モデルや構造制御機構との統合により，さらなる性能向上が期待される．

本研究における主研究と副次的研究の関係は，以下のように整理される：

- 主研究：領域の最適化（Where to transform）
- 副次的研究 1：エッジ構造の最適化（What structure to preserve）
- 副次的研究 2：周波数成分の最適化（What frequency components to enhance）

すなわち，本研究は領域と構造という二つの観点から画像生成を統合的に捉え，生成品質を向上させる枠組みとして位置付けられる．この統合的視点は，今後の画像生成技術の発展において重要な基盤となると考えられる．

今後は，まず大規模データセットを用いた評価を行い，提案手法の汎化性能をより詳細に検証する必要がある．特に，アニメ調画像は画風や属性の多様性が大きいため，異なるデータ分布に対しても安定した領域検出および局所変換が可能であるかを確認することが重要である．また，拡散モデルとのより高度な統合により，生成品質のさらなる向上を図ることも課題である．加えて，検出器と生成モデルを共同最適化する際に，生成品質がどのように検出領域の学習に影響を与えるのかを理論的に解析することで，本研究で導入した最適化枠組みの妥当性をより明確にする必要がある．最終的には，これらの検討を通じて，本研究の枠組みをイラスト制作支援やゲーム・映像制作などの実応用へ展開し，より汎用的かつ実用的な画像生成支援技術へと発展させることを目指す．

7 結論

本研究では，画像生成に適した領域抽出を目的とした新しい物体検出の枠組みを提案した．具体的には，

物体検出器と画像変換モデルを統合し，生成結果を評価信号として検出器の学習に反映する End-to-End 学習手法を構築した．これにより，従来の検出精度中心の最適化では得られなかった，生成に適した領域抽出が実現された．

さらに，本研究の過程で明らかとなった構造情報の重要性に基づき，副次的研究として，エッジ情報を利用した構造制御型画像修復手法，および双直交ウェーブレット変換を利用した周波数成分考慮型低照度画像強調手法を検討した．前者では，事前修復とエッジ洗練を組み合わせ，洗練されたエッジ情報を拡散モデルに条件として与えることで，構造的一貫性と視覚的自然さの両立を図った．後者では，重複型ウェーブレット変換と高周波成分補正を組み合わせることで，ブロック歪みを抑制しつつ，エッジやテクスチャを保持した自然な画像強調を実現した．

これらの成果は，アニメ調画像に限らず，画像編集，画像生成，画像修復，低照度画像強調を含む広範な画像処理・生成タスクにおいて有効であり，特に局所編集やインタラクティブ生成といった応用において大きな可能性を有している．

今後は，より高精度な生成モデルとの統合や大規模データセットによる検証を進めるとともに，本研究で提案した領域と構造の統合最適化という枠組みをさらに発展させることで，次世代の画像生成技術の基盤構築を目指す．

【参考文献】

[1] I. Goodfellow et al., “Generative Adversarial Nets,” *Proc. NeurIPS*, 2014.
[2] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” *Proc. NeurIPS*, 2020.
[3] W. Liu et al., “SSD: Single Shot MultiBox Detector,” *Proc. ECCV*, 2016.
[4] A. Bochkovskiy et al., “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv*, 2020.
[5] S. Ren et al., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE TPAMI*, 2017.
[6] A. Wang et al., “YOLOv10: Real-Time End-to-End Object Detection,” *Proc. NeurIPS*, 2024.
[7] J.-Y. Zhu et al., “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” *Proc. IEEE ICCV*, 2017.
[8] K. Nazeri et al., “EdgeConnect: Structure Guided Image Inpainting Using Edge Prediction,” *Proc. IEEE/CVF ICCV Workshop*, 2019.
[9] Y. Zeng et al., “Aggregated Contextual Transformations for High-Resolution Image Inpainting,” *IEEE TVCG*, 2023.
[10] A. Lugmayr et al., “RePaint: Inpainting Using Denoising Diffusion Probabilistic Models,” *Proc. IEEE/CVF CVPR*, 2022.
[11] H. Jiang et al., “Low-light Image Enhancement With Wavelet-Based Diffusion Models,” *ACM ToG*, 2023.
[12] G.X. Xu et al., “SNR-aware Low-light Image Enhancement,” *Proc. IEEE/CVF CVPR*, 2022.
[13] H.Y. Cai et al., “Retinexformer: One-Stage Retinex-Based Transformer for Low-light Image Enhancement,” *Proc. IEEE/CVF ICCV*, 2023.
[14] H.J. Hou et al., “Global Structure-aware Diffusion Process for Low-light Image Enhancement,” *Proc. NeurIPS*, 2023.
[15] D. Zheng et al., “Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model,” *Proc. IEEE/CVF CVPR*, 2024.
[16] W.B. Zou et al., “Wave-Mamba: Wavelet State Space Model for Ultra-High-Definition Low-light Image Enhancement,” *Proc. ACM MM*, 2024.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Overlapped Wavelet Diffusion for Low-Light Image Enhancement	IEICE Transactions on Information and Systems	2027 年 1 月掲載予定
アニメ調キャラクター画像における局所変換を考慮した End-to-End 物体検出の検討	2026 年電子情報通信学会 総合大会	2026 年 3 月
拡散モデルによる洗練エッジガイド付き画像修復	第 40 回信号処理シンポジウム	2025 年 11 月