

超解像を用いた道路反射鏡画像からの 車両検出の検証と評価

結城 光太郎^{1,a)} 栗 達^{1,b)} 小野 晋太郎^{1,c)}

概要: 本研究では、超解像技術を活用し道路反射鏡画像からの車両検出の有効性を検証した。車両画像(実像)、道路反射鏡画像(鏡像)を超解像技術を用いて高画質化し、その画像で車両検出を行い信頼度を評価した。結果として、鏡像は実像よりも信頼度が全体的に低くなり、鏡像において超解像技術を用いると入力画像サイズが小さい場合の信頼度が平均して18.8%向上するという結果が得られ、有効性が示された。

キーワード: 道路反射鏡, 超解像, 物体検出

Verification and Evaluation of Vehicle Detection from Traffic Mirror Images Using Super-Resolution

Abstract: In this study, we verified the effectiveness of vehicle detection from traffic mirror images using super-resolution technology. We enhanced the quality of vehicle images (real images) and traffic mirror images (mirrored images) using super-resolution and evaluated the confidence of vehicle detection on these images. As a result, the confidence of mirrored images was generally lower than that of real images. However, applying super-resolution to mirrored images improved the confidence by an average of 18.8% when the input image size was small, demonstrating the effectiveness of this approach.

Keywords: Traffic Mirror, Super-Resolution, Object Detection

1. はじめに

1.1 研究背景

交通事故の発生要因として、交差点が重要な危険箇所であることが知られている [1]。図1に示されるように、令和2年度の道路形状別交通死亡事故の発生件数は、交差点内が全体の約38.8%、交差点付近を含めると約50%を占めている。原因の多くは運転者の不注意によるものが大半であるが、道路環境が原因となる場合も少なくない。

また、日本では細い道が多く、建物や壁、植物などの障害物により枝道が死角となることがある。これらの死角を減らすため、日本では多くの道路反射鏡が交差点やT字路、急カーブ等に設置されている。道路反射鏡を通して、運転

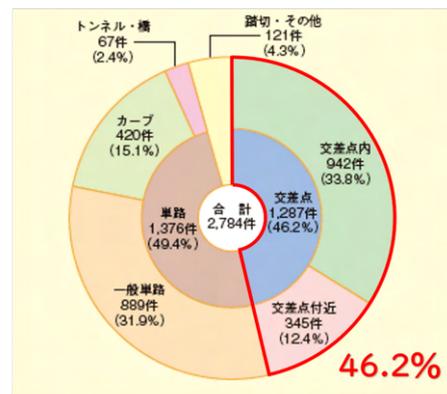


図1 道路形状別交通死亡事故発生件数(令和2年)

者は枝道の状況を確認し、歩行者や車両の動きを把握することができる。しかし、道路反射鏡内の状況を目視で確認するには、道路反射鏡に接近しなければ確認することが困難である。

¹ 福岡大学
Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka, 814-0001, Japan
a) t1211206@cis.fukuoka-u.ac.jp
b) lida@fukuoka-u.ac.jp
c) onoshin@fukuoka-u.ac.jp

1.2 研究目的

車載カメラの映像から道路反射鏡の状況を確認することが可能であれば、死角から車両や歩行者が接近している状況下で、目視よりも早く認識することが可能となり、運転の安全性向上に繋がる。

ここで問題となるのが、車載カメラで撮影された道路反射鏡の映像は低解像度であり、車両や歩行者を認識することが困難であるという点である。

そこで本研究では、超解像技術を用いた道路反射鏡画像からの車両検出の有効性を検証し、評価した。

2. 関連研究

道路反射鏡は、死角に優れた車両や歩行者の存在を確認するために重要な役割を果たしている。しかし、車載カメラで撮影された道路反射鏡画像は解像度が低いため、従来の物体検出技術では精度の向上に限界がある。Sato ら [2] は、低解像度の道路反射鏡画像における車両検出を目的とし、「Motion Feature Extraction Network」を提案している。この手法は、道路反射鏡内の画像から動き情報を抽出し、時間的な変化を活用して車両や歩行者の接近を予測するものである。また、実データ不足に対応するため、Unity と CycleGAN を用いて生成した合成データセットを活用することにより、学習精度の向上を図っている。道路反射鏡内の車両の検出精度を向上させるためには、低解像度画像に適応可能な手法が求められる。Sato らは、モーション特徴抽出と合成データセットの活用を組み合わせた手法を提案し、低解像度画像における車両検出の有効性を示した。

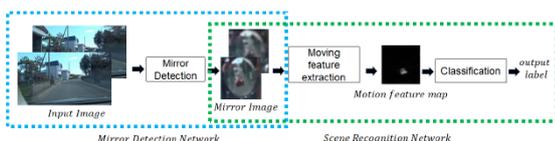


図 2 Motion Feature Extraction Network

図 2 のように提案されたネットワークは、3つのステップで構成されている。初めに、車載カメラで撮影された入力ビデオシーケンスから、RefineDet[3] を使用して道路反射鏡の領域を検出する。次に、連続する 9 枚の道路反射鏡領域画像から前景オブジェクトの動きの特徴を抽出する。この特徴抽出には、UNet[4] に基づく encoder-decoder network を使用し、3次元畳み込み (3d-conv) を採用している。最後に、motion feature map から接近するオブジェクトの有無を識別するため、ImageNet[5] で事前学習された ResNet-50[6] を使用する。これにより、低解像度のミラー画像からも効果的に接近物体を検出できる。

Sato らの手法は低解像度下での動きの特徴を抽出する手法がとられている。これに対し本研究では、道路反射鏡画像の解像度を直接改善する手法を検討する

3. 提案手法

超解像技術を活用し道路反射鏡画像からの車両検出の有効性を検証する為の提案手法を説明する。本研究では、SwinIR および Real-ESRGAN といった最新の超解像技術を活用し、道路反射鏡画像を高画質にすることで車両検出を行う手法の有効性を評価する事を目的としている。YOLOv9 の物体検出技術を用いて信頼度を評価し、それぞれの処理時間も評価対象とした。

3.1 超解像技術

本研究で使用した 2 つの超解像技術について説明する。

3.1.1 SwinIR

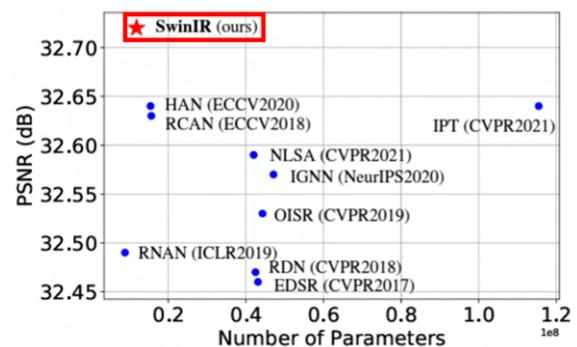


図 3 Set5[7] における画像超解像 (× 4) のためのさまざまな手法における、PSNR 結果と総パラメータ数の比較

Swin Transformer という transformer ベースのモデルが使用されており、従来の CNN ベースモデル手法よりも優れた性能を発揮する。低解像度の画像、ノイズが含まれている画像、圧縮された画像などを得意とする。高解像度の画像を扱う場合、処理時間が増大してしまう欠点がある。

図 3 の、Set5 は画像超解像 (SR) タスクでよく使用される小規模なベンチマークデータセットの 1 つである。このデータセットは超解像アルゴリズムの性能を評価するために広く利用されている。Set5 を使用し、超解像モデルのパフォーマンスを評価する際には、主に PSNR (ピーク信号対雑音比) や SSIM (構造類似度指数) などの指標が使用される。これらの指標により、復元画像の品質が元の高解像度画像にどれだけ近いかが評価される。

図 3 より SwinIR は他のモデルと比較し、少ないパラメータでより高い PSNR (ピーク信号対雑音比) を実現していることが分かる。PSNR の高さは、復元画像や圧縮後の画像が元の画像に近いことを示している。

SwinIR[8] は浅い特徴から深い特徴までを段階的に抽出し、劣化した画像を高品質な画像に復元している (図 4)。初めに、浅い特徴抽出を行い入力画像から基本的な特徴を抽出する。ここでは主に畳み込み層 (Conv 層) が使用され、エッジ、テクスチャなどの画像の低レベルな情報が

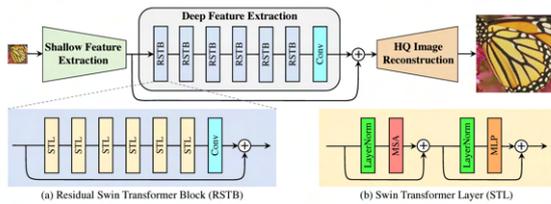


図 4 画像復元のために提案された SwinIR のアーキテクチャ

取り込まれる。

次に、深い特徴抽出が行われる。Residual Swin Transformer Block (RSTB) が複数並べられており、これが画像の中から大量の特徴を抽出する。RSTB の内部には、複数の Swin Transformer Layer (STL) が含まれており、STL が画像の異なる部分を注意し特徴を取り出す。Residual Swin Transformer Block(RSTV) は出力されたものが再び入力として加えられる残差接続であるため、特徴抽出の精度を高めながら、情報の損失を保護している。

最後に再び畳み込み層を使い、最終的な画像を生成するという仕組みである。

3.1.2 Real-ESRGAN



図 5 実際の画像における Bicubic アップサンプリング, ESRGAN[10], RealSR[11], Real-ESRGAN の結果比較. 合成データで訓練された Real-ESRGAN モデルは、一般的な実世界の画像に対し、煩わしいアーティファクトを除去しながら、詳細を強調することが可能である。

Real-ESRGAN[9] (Real-Enhanced Super-Resolution Generative Adversarial Networks) は、画像超解像技術の一つであり、特にリアルな画像の質感や細部を保持しながら解像度を向上させることを目指したモデルである。

従来の超解像モデルではノイズの少ない画像を対象としていたのに対し、「高次の劣化」と述べられている写真やビデオに含まれるノイズや劣化を考慮した設計になっているため、現実のノイズや劣化のある画像に対して優れた性能をもつ。

図 5 より、Real-ESRGAN は他のモデルと比べ鮮明な画像になっていることが分かる。

3.2 物体検出技術

本研究で利用した物体検出技術について説明する



図 6 物体検出後の出力画像

3.2.1 YOLOv9

物体検出アルゴリズムである YOLO(You Only Look Once) は、名の通り入力画像を一度だけ見る (Look Once) ことで物体を検出するという特徴がある。従来のアルゴリズムは、画像を複数回処理する必要があったが、YOLO は画像全体を一度に処理し、物体の位置とクラスを予測できるため、高速かつ効率的である。リアルタイムでの物体検出が可能である。

図 6 のように、0~1 の間で信頼度が表されており信頼度が 1 に近いほど「物体」であることを表している。

YOLOv9[12] では、従来のバージョンよりもモデルが軽量化され、リアルタイム処理の性能と計算効率が上がっている。より小さな物体の検出能力や、複雑な背景における物体の認識が可能である。

YOLOv9 では、programmable gradient information (PGI) という概念が提案されており、既存の手法が入力データがレイヤーごとの特徴抽出と空間変換を受けると、大量の情報が失われるという事実を無視しているのに対し、PGI は目的関数を計算するための目的タスクの完全な入力情報を提供することができ、ネットワークの重みを更新するための信頼性の高い勾配情報を得ることができる。

4. システムの概要

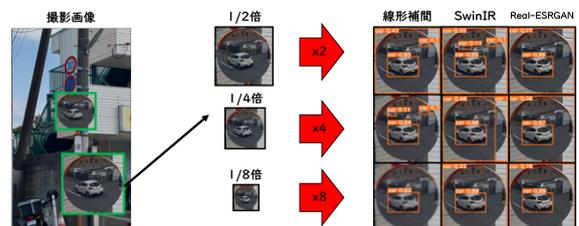


図 7 提案手法

本研究では、線形補間, SwinIR, Real-ESRGAN それぞれで処理を施した道路反射鏡に映る車両画像の物体検出を行い、信頼度および処理速度を比較評価するため、以下の手法を提案する (図 7)。

初めに、車両画像と道路反射鏡画像を iPhone で撮影する。サイズは、1180 × 2080(px) であり、撮影画像から車両画像 (実像) および道路反射鏡画像 (鏡像) をそれぞれ抽出し、これらを正解画像 (original 画像) として設定する。

次に、抽出した画像に対して 1/2 倍、1/4 倍、1/8 倍とリサイズ処理を実施する。リサイズ画像を入力画像とし、2 倍、4 倍、8 倍で超解像処理を行う。この際、それぞれのアルゴリズムで処理時間を記録する。2 倍、4 倍、8 倍の処理を各 5 回ずつ実行し、倍率ごとの平均実行時間を記録。また、歪み補正を考慮する場合、正確な 3 次元位置画像が必要であるため本実験では省略する。

最後に、YOLOv9 による検出を行い信頼度を比較・評価する。本実験では、車両の検出が目的であるため車両のみ検出するようにクラスを指定する。線形補間、SwinIR、Real-ESRGAN の 3 つのアルゴリズムで処理した画像に対し、YOLOv9 で物体検出を行う。各倍率ごとで、実像 6 枚の平均信頼度と鏡像 11 枚の平均信頼度をそれぞれ記録する。

以上の手順で、線形補間、SwinIR、および Real-ESRGAN の超解像手法について、それぞれの性能および有効性を詳細に評価した。

5. 実験結果・評価

5.1 実像の結果・評価



図 8 YOLOv9 による検出結果: 実像 x2

図 8 より信頼度の結果は、線形補間は 83%~96%, SwinIR は 82%~96%, Real-ESRGAN は 93%~96% であり、Real-ESRGAN の信頼度が高いことが分かる。

線形補間と超解像での比較の結果は、信頼度の差が 10% 未満とほとんど差はなかった。また、視覚的には超解像の方が高画質であった。original 画像と比較すると、同程度の信頼度で検出されていることが分かる。

誤認識は見受けられなかった。

図 9 より信頼度の結果は、線形補間は 86%~95%, SwinIR は 81%~96%, Real-ESRGAN は 79%~96% であり線形補間の信頼度が高いことが分かる。

線形補間と超解像での比較の結果は、信頼度の差が 10% 未満でありほとんど差はなかった。また、視覚的には超解像の方が高画質であった。original 画像と比較すると、同



図 9 YOLOv9 による検出結果: 実像 x4

程度の信頼度で検出されていることが分かる。誤認識は見受けられなかった。



図 10 YOLOv9 による検出結果: 実像 x8

図 10 より信頼度の結果は、線形補間は 0%~58%, SwinIR は 58%~91%, Real-ESRGAN は 0%~92% であり、SwinIR の信頼度が高いことが分かる。

線形補間では画像のぼやけが酷く、信頼度も未検出 (0%)~58% と低い信頼度であった。これに対し、超解像では線形補間に比べ画質がクリアで、92% と高い信頼度で物体検出されていた。これは original 画像と同程度の信頼度である。特に、SwinIR に関しては未検出がなく 58%~91% と高い信頼度で検出された。

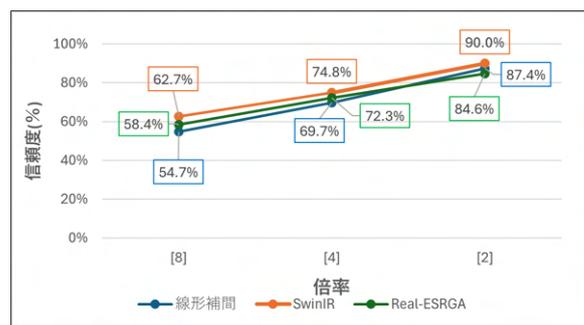


図 11 線形補間、SwinIR、Real-ESRGAN の 3 つのアルゴリズムでの実像における各倍率ごとの信頼度

図 11 は線形補間、SwinIR、Real-ESRGAN の 3 つのアルゴリズムにおいて各倍率ごとで、実像 11 枚の平均信頼度を記録したグラフである。

倍率が小さくなる、つまり入力画像サイズが大きくなるにつれ、信頼度が高くなる傾向がある。また、3 つのアルゴリズムの中で SwinIR がどの倍率においても信頼度が高

く、線形補間との信頼度の差は8%と小さかった。

5.2 鏡像の結果・評価

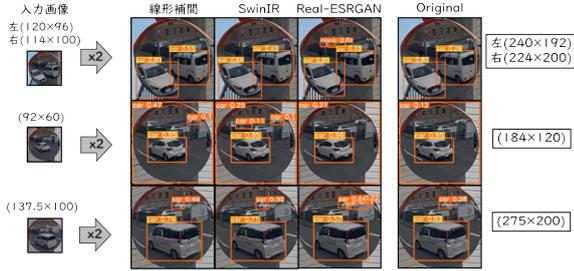


図 12 YOLOv9 による検出結果: 鏡像 x2

図 12 から分かる通り信頼度の結果は、線形補間は 53%~95%, SwinIR は 43%~95%, Real-ESRGAN は 33%~96% であり、Real-ESRGAN の信頼度が高いことが分かる。

軽ワゴン車の信頼度において、20%ほどの差が出た。視覚的な差はほとんどない。original 画像と比較すると、同程度の信頼度で検出されていることが分かる。

また、道路反射鏡が車として認識されてしまう誤認識が確認される。



図 13 YOLOv9 による検出結果: 鏡像 x4

図 13 から分かる通り信頼度の結果は、線形補間は 29%~94%, SwinIR は 55%~96%, Real-ESRGAN は 66%~97% であり、Real-ESRGAN の信頼度が高いことが分かる。

線形補間では軽ワゴン車の信頼度が 29% であるのに対し、超解像では信頼度が 55%~66% と高くなっていることが分かる。

また、視覚的にも超解像の方が高画質である。original 画像と比較すると、同程度の信頼度で検出されていることが分かる。

超解像後の道路反射鏡が車両として検出される誤認識が確認される。

図 14 から分かる通り信頼度の結果は、線形補間は 0%~62%, SwinIR は 0%~89%, Real-ESRGAN は 36%~89% であり、Real-ESRGAN の信頼度が高いことが分かる。

線形補間では画像のぼやけが酷く、信頼度も未検出(0%)~62%と低い信頼度であった。これに対し、超解像では線形補間に比べ画質がクリアで、92%と高い信頼度で物体検出



図 14 YOLOv9 による検出結果: 鏡像 x8

されていた。これは original 画像と同程度の信頼度である。特に、SwinIR に関しては未検出がなく 58%~91% と高い信頼度で検出された。

超解像後の道路反射鏡が車両として検出される誤認識が確認される。

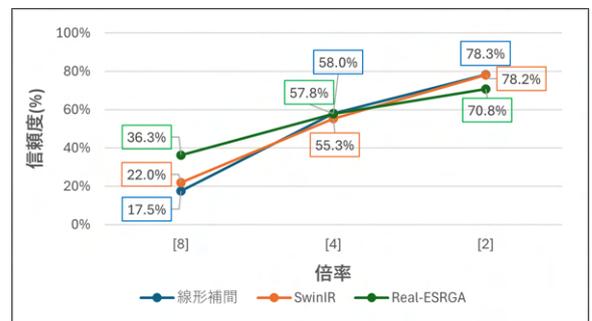


図 15 線形補間, SwinIR, Real-ESRGAN の 3 つのアルゴリズムでの鏡像における各倍率ごとの信頼度

図 15 は線形補間, SwinIR, Real-ESRGAN の 3 つのアルゴリズムにおいて各倍率ごとで、鏡像 6 枚の平均信頼度を記録したグラフである。

倍率が小さくなる、つまり入力画像サイズが大きくなるにつれ、信頼度が高くなる傾向があった。また、[x8] では Real-ESRGAN の信頼度が高く、線形補間と Real-ESRGAN との信頼度の差は 18.8% と大きく開いた。また、[x8] になると、線形補間は信頼度が著しく減少することが分かる。

5.3 実像における信頼度と鏡像における信頼度の比較

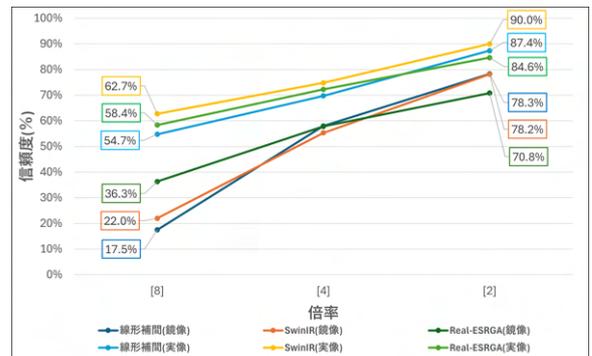


図 16 線形補間, SwinIR, Real-ESRGAN の 3 つのアルゴリズムでの実像・鏡像における各倍率ごとの信頼度

図 16 は図 11 と図 15 を比較したグラフである。

図 16 から、実像の方が鏡像よりも全体的に信頼度が高くなっていることが分かる。また、実像・鏡像共に [x8] では超解像の方が線形補間より信頼度が高いことが分かる。

5.4 同一画像・異なる倍率での比較



図 17 YOLOv9 による検出結果：同一画像・異なる倍率

図 17 は同一画像・異なる倍率での比較の図である。信頼度の結果は、[x2] では 57%~74%，[x4] では 74%~80%，[x8] では 0%~65% である。同じ画像で比較することにより倍率ごとの信頼度の差が見えてくる。

[x2]，[x4] では、どのアルゴリズムでも 57%~80% と高い信頼度で検出されている。それに対し、[x8] では未検出や信頼度が 55% と低い結果が出ている。視覚的にも入力画像が大きい方が処理後の画質は鮮明になっていることが分かる。

5.5 超解像処理時間の比較



図 18 線形補間，SwinIR，Real-ESRGAN の 3 つのアルゴリズムでの処理時間の比較

図 18 は、線形補間，SwinIR，Real-ESRGAN の 3 つのアルゴリズムの処理時間を記録したグラフである。

線形補間での処理時間は 0 秒に近く、ほとんど処理時間がかからないことが分かる。SwinIR は倍率が小さくなる、つまり入力画像が大きい方くなるにつれ、処理時間が増加している。Real-ESRGAN は、入力画像の倍率にかかわらず、処理時間は 5~6 秒の間と一定であることが分かる。

6. 結論

6.1 考察

[x2][x4] では、実像、鏡像共に線形補間、SwinIR、Real-ESRGAN の 3 つのアルゴリズムの処理でほとんどの画像が original 画像との信頼度のが差が 5% 以内で検出されていた。この結果より、入力画像が大きいサイズ ([x2]，[x4]) の場合、短い処理時間 (1 秒未満) で実行できる線形補間を行うのが良いと考察できる。(図 8, 9, 12, 13, 18 より)

[x8] では、実像、鏡像共に線形補間画像に比べ超解像処理された画像の方が 30%~40% 程度高い信頼度で物体検出されていた。この結果より、入力画像が小さいサイズ ([x8]) の場合、超解像処理が有効であると考察できる。また、処理時間を考慮すると、SwinIR を使用すれば、1 秒程度の短い処理時間で高い信頼度を得ることができると考察できる。(図 10, 14, 16, 18 より)

実像と鏡像の信頼度の比較から、実像の方が鏡像に比べ全体的に高い信頼度で物体検出されていることが分かる。この結果から、鏡像は道路反射鏡越しの車両画像であるため、車両全体が写っていない事や、道路反射鏡の曲率の影響、また、道路反射鏡内の明るさや光の反射等が物体検出に影響していると考えられる。(図 16)

これらの考察を踏まえ、超解像処理を行うことにより遠くの道路反射鏡の車両の認識が可能なり、超解像技術を用いた道路反射鏡画像からの車両検出における、手法ごとの信頼度の違いを検証し、有効性を示すことができたとと言える。

6.2 課題

超解像処理後の画像でも物体検出されていないものや、道路反射鏡が車両として誤認識されてしまう場合があった。また、超解像処理に時間がかかってしまい、リアルタイム処理を実現するには厳しい結果となった。

6.3 今後の展望

車載カメラ映像で超解像処理を行い、道路反射鏡内の車両をリアルタイム検出することで、運転者に危険を知らせるシステムを構築したいと考えている。そのために、超解像処理の高速化や検出精度の向上など、アルゴリズムの構築を追求していきたい。

謝辞 本研究の一部は JSPS 科研費 JP21H01457, JP21K03962, および NEDO SIP (スマートモビリティプラットフォームの構築) の支援を受けた。

参考文献

- [1] 内閣府, 第1編 陸上交通 第1部 道路交通 第1章 道路交通事故の動向 https://www8.cao.go.jp/koutu/taisaku/r03kou_haku/zenbun/genkyo/h1/h1b1s1.html, (最終アクセス日: 2025年1月15日)
- [2] Rikiya Sato, Kazunori Onoguchi, Scene Recognition for Blind Spot via Traffic Mirror and In-Vehicle Camera, 2022
- [3] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li. Single-shot refinement neural network for object detection. In CVPR, 2018.
- [4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015.
- [5] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deep Residual Learning for Image Recognition, 2015
- [7] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In British Machine Vision Conference, pages 135.1-135.10, 2012.
(最終アクセス日: 2025年1月15日)
- [8] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, Radu Timofte SwinIR: Image Restoration Using Swin Transformer
- [9] Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data
- [10] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In ECCVW, 2018.
- [11] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In CVPRW, 2020.
- [12] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information