

修士論文要旨 (2022 年度)

YOLOv7-tiny ベースの軽量で高精度な自動ナンバープレート認識 A Light-Weight and High-Accuracy Automatic License Plates Recognition Based on YOLOv7-tiny

電気電子情報通信工学専攻

Department of Electrical, Electronic, and Communication Engineering

塩澤 馨

Kyo Shiozawa

1 はじめに

ALPR(自動ナンバープレート認識) システムはこれまで、駐車場における契約車両の判定、交通量調査、商業施設などでのマーケティングなどで利用されてきた。また、N システムでは犯罪車両の追跡にも利用されている。

ALPR システムはナンバープレートの検出、文字認識の 2 つもしくは、高解像度画像を使用する場合に車両検出を追加した 3 つのタスクでナンバープレートの認識を行う。近年のハードウェアの発展と共にディープラーニングも大きく発展した。また、ALPR システムもこれまでの事前に設定したパターン認識を用いる手法はディープラーニングを取り入れることで認識の精度が大きく向上した。ALPR システムのタスクにディープラーニングを適用することでこれまで独立して行ってきた 2 つのタスクを同時に 1 つのネットワークで実行できるようになった。[3] これにより、一度の検出でナンバープレートの認識を行うことができ認識速度が向上した。

近年、煽り運転などの危険運転に関する報道が増えたことでドライブレコーダーの普及が進んだ。これにより車両を撮影するカメラの台数が増えた。将来的にドライブレコーダーなどで撮影した画像をもとに高速でナンバープレートの認識を行うことで得られる利益は大きい。例えば、これまで N システムが行ってきた静的な追跡を動的に行うことができるようになると考えている。そのため軽量で高精度な

ALPR システムが必要になる。

2 提案手法

本研究で提案する ALPR システムは 1 度にナンバープレートの検出と文字の認識を行う。検出を行うネットワークから得た物体の情報を組み立ててモジュールでナンバープレートとして組み立てを行いナンバープレートの認識を行う。YOLOv7-tiny[2] をベースとした高い精度で検出を行えるネットワークを組み立ててモジュールと組み合わせて軽量で高精度な ALPR システムを構成する。

2.1 組み立てモジュール

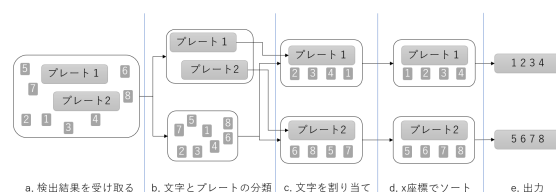


図 1: 組み立てモジュールの処理の流れ

組み立てモジュールの処理の流れを図 1 に示す。まず、a の処理でネットワークから検出された物体の情報を受け取る。その後、b の処理で検出された物体をプレートと文字に分類する。この際に設定された信頼度スコアの閾値以下の物体は破棄する。その後各文字がどのナンバープレートに属するか判定を行い振り分けを行う。d で振り分けられた文字を x 座標で sort し、e で出力を行う。

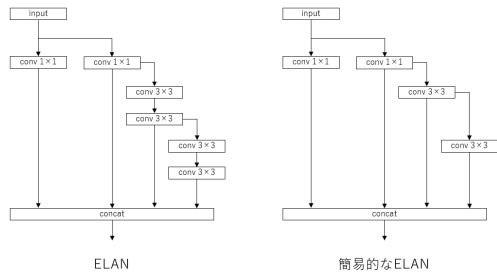


図 2: ELAN と簡易的な ELAN

2.2 検出部の拡張

YOLOv7-tiny の検出速度を保ちつつネットワークが高い精度で物体の認識を行うために検出部に簡易的な ELAN の追加を行った。ELAN は入力のチャンネル数を分割しそれぞれで畳み込みを行うことで高速な処理が可能になる。簡易的な ELAN は ELAN から畳み込み層 2 つ除きより軽量になっている。これをネットワークのすべてのスケールの検出部の適用する。

2.3 活性化関数の修正

ニューラルネットワークの強味の一つは非線形関数を表現できることでありそれを実現しているのが活性化関数である。代表的な活性化関数にシグモイド関数や ReLU 関数がある。YOLOv7-tiny は Leaky ReLU 関数を用いる。提案手法のネットワークでは活性化関数を Leaky ReLU 関数から Swish 関数に修正を行う。Swish 関数は下記の数式で表される。

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

$$f(x) = x \cdot \sigma(\beta x) = \frac{x}{1 + e^{-\beta x}} \quad (2)$$

Swish 関数は入力値 x が負の領域では出力値はわずかに負の値をとるがほとんど 0 になり、正の領域では 0 に近い値で多少減少するがほとんど入力値 x と同じ値を出力する。

2.4 データ拡張の追加と削除

物体検出の精度を向上させる方法の 1 つにデータ拡張がある。データセットの作成は非常にコストのかかる作業であり、特に物体検出でのデータセット

の作成は画像中のどの部分に物体がありそれがどのクラスに属するのかを指定する必要がありよりコストがかかる作業となる。作成したデータセットに加工を行い、データセットの拡張を行うことでデータセットの作成のコストをかせずにネットワークが十分に学習可能な量のデータセットを確保する。

本研究で使用するディープラーニングのフレームワークである Darknet で実装された YOLOv7-tiny では、デフォルトで彩度、明度、色相、回転によるデータ拡張が使用されている。これに加えて mosaic データ拡張、ネガティブポジティブ反転によるデータ拡張を追加する。形が統一されていない動物とは異なりナンバープレートは形が統一されている。よって回転によるデータ拡張により実際には存在しないナンバープレートの特徴を学習し、認識精度を悪化させる。そのため本研究の提案手法のネットワークにはデフォルトで使用される回転によるデータ拡張は使用しない。また、ナンバープレートと他の物体が重なった場合に検出することはできないため考慮しない。よって複数の画像を重ね合わせる MixUp や CutMix によるデータ拡張を使用しない。

2.5 ネットワークの拡張

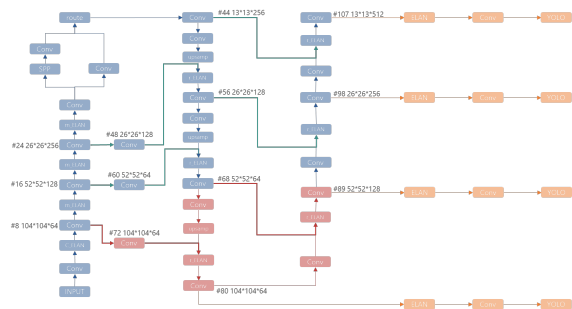


図 3: 提案手法のネットワーク

YOLOv7-tiny は $13 \times 13, 26 \times 26, 52 \times 52$ の 3 層で物体の検出を行う。入力画像の畳み込み処理を行い、物体の検出に使用する特徴マップの作成を行う。作成された特徴マップを各スケールの特徴マップと共有することで小さいスケールのもつ詳細特徴マップを大きいスケールの特徴マップでも使用することが可能になり小さい物体の検出精度の向上している。

しかし、入力画像において小さい物体になりやすいナンバープレートは畳み込み処理の過程で情報を失いやすくなる。そのため提案手法のネットワークは104×104の検出部を追加した4層での検出を行う。

3 評価実験

後述する自作したデータセットの訓練データと検証データを用いて学習を行う。テストデータでナンバープレートの認識の精度の検証を行い、ドライブレコーダーから撮影した1分の動画5本を用いて速度の比較を行う。

3.1 データセット

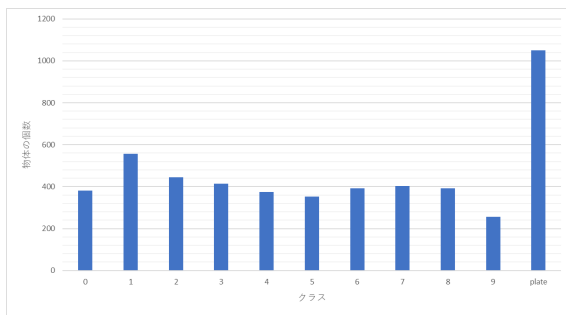


図4: 自作データセットの各クラスの分布

公開された国内のナンバープレートのデータセットがなかったため、データセットを自作した。東京都を中心に走行中の車両のドライブレコーダーで撮影した映像を画像に切り出し416×416のサイズにトリミングを行った。1052枚の画像データにアノテーションを行い、訓練データ、検証データ、テストデータにそれぞれ736枚、210枚、106枚に分配した。plate(ナンバープレート)、と0から9の11クラスでアノテーションを行った。自作データセットに含まれるアノテーションされた各クラスの分布は表4に示す。

3.2 認識精度の評価

3.2.1 認識精度の評価の手法

自作したデータセットのテストデータを用いて提案手法のALPRシステムの認識精度の検証を行う。評価の軸としてmAPと各信頼度スコアの閾値でのALPRシステムの認識の正答率を使用する。mAPとは各クラスのAPの平均である。APとは物体検

出に使われる評価指標であり、検出された物体の総数に含まれる正解データの割合を表すPrecisionを縦軸にとり、すべての正解データの数に含まれる検出された正解データの割合を表すRecallを横軸にとり、代表点を用いて積分を行い算出される。信頼度スコアはネットワークがどの程度の信頼度でその物体を検出したかを表す。

3.2.2 認識精度の評価の結果

表1: ネットワークの精度の比較

ネットワーク	mAP	@90	@85	@80	@75	@70
YOLOv4	94.85	59.43	62.26	62.26	58.49	51.89
YOLOv4-t	93.27	57.55	59.43	58.49	63.21	65.09
YOLOv7	97.43	0.00	4.72	33.02	44.34	62.26
YOLOv7-t	96.18	5.66	24.53	43.40	59.43	66.04
YOLOv7x	97.54	0.00	10.38	50.00	70.75	83.02
提案手法	99.03	0.00	44.43	66.92	88.68	99.06

結果を表1に示す。tの表記はそれぞれYOLOv4-tiny-3l[1], YOLOv7-tinyである。mAPの比較では提案手法のネットワークが最も高い値となった。各信頼度スコアでの認識の正答率で比較を行うと高い閾値ではYOLOv4を用いたALPRシステムが最も精度が高かったが、閾値が80以下では提案手法のネットワークを用いたALPRシステムが最も精度が高かった。また、YOLOv4系を用いたALPRシステムでは閾値によらず一定の正答率となるのに対して、提案手法を含むYOLOv7系を用いたALPRシステムでは閾値を下げる正答率が向上しYOLOv4系を用いたALPRシステムより精度が高くなった。

3.3 検出速度の評価

3.3.1 検出速度の評価の手法

ドライブレコーダーから撮影した1分の動画5本を用いて提案手法のネットワークの検出速度の比較を行った。1分間の動画再生中に各ネットワークで検出を行い、FPSを計測する。

3.3.2 検出速度の評価の結果

表2: ネットワークの速度の比較

ネットワーク	FPS1	FPS2	FPS3	FPS4	FPS5	ave FPS
YOLOv4	34.0	34.5	35.0	35.5	35.5	34.9
YOLOv4-t	64.0	64.0	63.7	64.1	64.0	64.0
YOLOv7	57.6	57.3	57.7	59.0	58.5	58.0
YOLOv7-t	64.0	63.7	64.0	63.7	64.0	63.9
YOLOv7x	32.0	32.0	32.0	32.0	32.0	32.0
提案手法	64.1	64.0	64.0	64.0	64.0	64.0

結果を表 2 に示す。

YOLOv4, YOLOv7, YOLOv7x と比較して提案手法のネットワークの検出速度は高速であり, YOLOv4, YOLOv7 の軽量モデルである YOLOv4-tiny-3l, YOLOv7-tiny と比較しても同程度の検出速度となった。

4 考察



図 5: 提案手法でのナンバープレートの認識結果の例

提案手法でのナンバープレートの認識結果の例を図 5 に示す。図 5 の画像は他のネットワークを用いた ALPR システムでは正しく認識することができなかったが、提案手法では正しく認識を行うことができた。(a) や (b) の車体とナンバープレートの色が似ている場合にも境界を正しく認識できていることがわかる。また (b) や (c) の緑や黄色など、少数になる配色のナンバープレートも正しく認識が行えている。提案手法のネットワークは検出部の追加を行い、追加した検出部のスケールは検出部の中で最大の 104×104 で、これは入力画像の $\frac{1}{4}$ のスケールである。これによりナンバープレートのような小さい物体でも畳み込みによるダウンスケーリングで情報を失うことがなく、畳み込みで抽出した詳細な情報と組み合わせて検出を行うことができていると考えられる。また、データセットに対して少数になる緑や黄色のナンバープレートは、ネガティブポジティブ反転によるデータ拡張により十分な量の学習データを確保することができるようになり認識が可能になったと考えられる。

ネットワークの検出部の拡張や検出部の追加を行ったが提案手法のネットワークのベースとなった YOLOv7-tiny と同程度の検出速度を維持した。拡張には主に簡易的な ELAN を使用しており、チャネ

ルの分割を行いながら畳み込みを行うことで検出速度を落とさずに精度の向上が可能になっている。

5 まとめ

YOLOv7-tiny をベースとして検出部の拡張、活性化関数の修正、データ拡張の追加と削除、ネットワークの拡張を行った。信頼度スコアの閾値が 70 で認識の正答率が 99.06% と高精度で検出を行うことができた。本研究ではデータセットを自作したため数字のみの認識までとしたが、日本のナンバープレートは数字以外にもひらがなや都市名を含みサイズは数字よりも小さい。また、海外で多い横一列に文字が並ぶナンバープレートではなく縦横の配置を考慮しなければならないため認識はより難しくなると考えられる。そのためすべての文字を含むデータセットでの検証を行い、必要であれば修正を行う必要がある。認識が困難であった場合、高解像度の画像からナンバープレートの検出しトリミングを行ってからナンバープレートの認識を行う 2 ステップの認識にする必要があるがその場合にも軽量で高精度で検出を行える提案手法のネットワークは貢献できる。

参考文献

- [1] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv, 2004, 10934
- [2] Wang, Chien-Yao and Bochkovskiy, Alexey and Liao, Hong-Yuan Mark, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”, 2022, arXiv:2207.02696.
- [3] Q. Huang, Z. Cai and T. Lan, “A Single Neural Network for Mixed Style License Plate Detection and Recognition,” in IEEE Access, vol. 9, pp. 21777-21785, 2021, doi: 10.1109/ACCESS.2021.3055243.