

Few-shot 物体カウントで複数種類の物体をカウントする際の過検出抑制技術

Method with Simple Loss Function Reducing False Positives
to Predict Number of New Objects in Images for Few-Shot Object Counting Technology

少数の正解教示で新規物体の個数を推定する Few-Shot 物体カウントにロス関数を導入

画像中の物体の個数を数える物体カウント技術の中で、近年、Few-shot 物体カウントが注目されています。学習データに含まれない新規物体も、推論時に少数の正解教示を与えるだけでカウントできるため、現場での再学習が不要で、実用上有用です。しかし、従来手法は、画像中に複数のカテゴリーの物体が含まれる場合、過検出することがありました。そこで、対象の物体が写っていない負例データを学習データに追加するシンプルな学習手法を提案しました。検証の結果、従来手法に比べて約 40 % の過検出低減を確認しました。

はじめに

物体カウントとは画像中の物体の個数を数える技術であり、防犯映像中の群衆の人数計測などに応用できます。一方、近年、推論時に少数の正解教示を与えることで新規カテゴリーの物体の個数を数えられる Few-shot 物体カウントが研究されています。人物など特定のカテゴリーに特化した従来のモデルと異なり、再学習を行うことなく新規カテゴリーに対応できる利点があります。この技術により、工場や物流現場など商品の入れ替えが頻繁に発生する場合でも、即座に対応できるようになります。

しかし、従来手法では、画像中に複数のカテゴリーの物体が写っている場合、対象外の物体も数えてしまう過検出の傾向がありました。いちごとぶどうが写っている画像（図 1 (a)）に従来手法⁽¹⁾を適用した結果を、図 1 (b) に示します。入力画像中の数えたい物体（ぶどう）の正解教示（Exemplar）を与えたとき、物体の個数を表す密度マップを出力した結果です。対象のぶどうだけでなく、対象外のいちごも数える過検出が示されました。

そこで、複数のカテゴリーの物体が写っていても、カテゴリーごとに個数を数えられるように、Few-shot 物体カウントの過検出を低減する手法を提案しました。

過検出低減のためのロス関数の提案

Few-shot 物体カウントの深層学習モデルは、入力として

画像と Exemplar を与えたとき、物体の存在する領域を表す密度マップを出力します（図 2）。モデルはあらかじめ大量の教師データを用いて学習し、推論時には数えたい物体の Exemplar を少数与えるだけで、再学習は不要です。学習時のロス関数は式(1)の最小二乗誤差 L を用います。

$$L = \sum_{i=1}^N \|y_i - f(x_i, b_i)\|^2 \quad (1)$$

ただし、 N は学習データの数、 x_i 、 b_i 、 y_i 、 $f(x_i, b_i)$ はそれぞれ i 番目のサンプルの画像、Exemplar（物体の画像中の矩形（くけい）座標）、正解の密度マップ、学習対象のモデル f が出力する密度マップです。しかし、従来手法で学習データとして広く用いられている公開データは、数えたい物体が必ず含まれている画像と密度マップと Exemplar の正解教示であり、数えたい物体が写っていない画像は含まれません。このようなデータで学習すると、モデルは物体を過検出しやすくなる懸念があります。

そこで、提案手法では学習データに数えたい物体が写っていない別の画像を追加し、式(2)のようなロス関数で学習します。

$$L = \sum_{i=1}^N \|y_i - f(x_i, b_i)\|^2 + \sum_{i=1}^N \|f(x'_i, b_i)\|^2 \quad (2)$$

ただし、 x'_i は b_i によって与えられる物体カテゴリーを含まない、 x_i とは別の画像です。以下、 x_i と x'_i をそれぞ



図 1. 従来手法と提案手法によるぶどう個数の推定結果（MAE）の比較
提案手法は、従来手法の過検出を低減します。

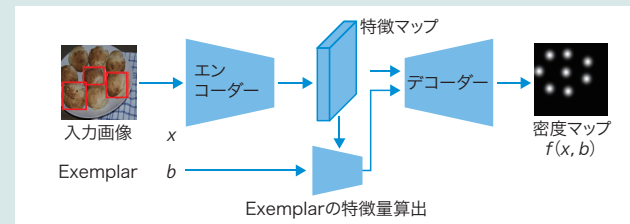


図 2. Few-shot 物体カウントの深層学習モデル

モデルは、入力画像と Exemplar から、密度マップを出力します。

れ正例データ、負例データと呼びます。第 2 項が画像に Exemplar の物体カテゴリーが含まれないことに対応し、この項によって過検出の低減が期待されます。

学習時の処理を述べます。入力した学習データに対してエンコーダーが出力した特徴マップと Exemplar から得られる特徴マップに基づいて、正解密度マップをデコーダーによって回帰します。これに対して、負例データを入力する処理を追加しました。追加した負例データには数えたい物体が写っていない画像を選別しているので、出力すべき密度マップはゼロとなり、これが式(2)の第 2 項に対応します。

提案手法は学習時に負例データを追加するシンプルなものであり、従来手法のモデルに容易に追加できます。

提案手法による過検出低減効果の確認

提案手法を従来手法⁽¹⁾に適用し、過検出の低減効果を確認しました。使用データは公開データセット FSC147⁽²⁾です（カテゴリー数 147、物体の平均個数 56）。

ここでは二つの実験を行いました。実験 1 は Exemplar として少数（3 個）の正解教示だけを与えたときの個数の推定であり、これは従来手法と同一です。一方、実験 2 は Exemplar として与えた物体とは異なる物体が写っている画像を入力しました。個数の正解は 0 であり、この実験 2 によって過検出の低減を確認します。図 3 に、実験 1 及び 2 における正解と推定結果の個数の平均絶対誤差（MAE）を比較しました。実験 1 では、提案手法の MAE は 15.94 と従

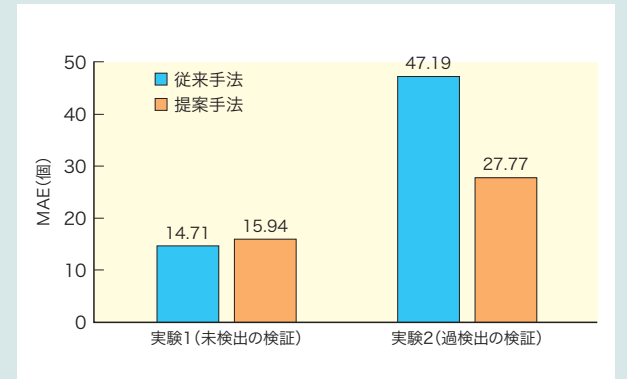


図 3. 提案手法による誤検出低減効果の検証結果

物体の個数を推定する性能を維持しながら、過検出を約 40 % 削減しました。

来手法の 14.71 と同程度であり、数えたい物体の個数を正しく推定する性能も維持できました。実験 2 では、従来手法の MAE が 47.19 でしたが、提案手法では 27.77 に改善し、過検出を約 40 % 削減できました。

図 1 の例でも、提案手法による過検出の低減が確認できました（図 1 (c)）。

おわりに

今後は、更なる誤差低減や推論処理の高速化などの改良を進め、実用化を目指します。

文 献

- (1) Liu, C. et al. "CounTR : Transformer-based generalised visual counting". British Machine Vision Conference (BMVC) 2022, London, 2022-11. <https://bmvc2022.mpi-inf.mpg.de/0370.pdf>, (accessed 2024-12-20).
- (2) Ranjan, V. et al. "Learning to count everything". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) . 2021, p.3394-3403. <https://openaccess.thecvf.com/content/CVPR2021/papers/Ranjan_Learning_To_Count_Everything_CVPR_2021_paper.pdf>, (accessed 2024-12-20).

齊藤 廣大

総合研究所 AI デジタル R&D センター コラボレイティブ AI 研究部
博士（情報理工学）、情報処理学会会員