

順序の一貫性を保った順序回帰手法の撮影年推定への適用

芦田 聖太[†] AdamJatowt[†] 吉川 正俊[†]

[†] 京都大学大学院社会情報学専攻吉川研究室 〒606-8501 京都府京都市左京区吉田本町
E-mail: †{shota.ashida,yoshikawa}@db.soc.i.kyoto-u.ac.jp, ††adam@dl.kuis.kyoto-u.ac.jp

あらまし 現在大量の画像がインターネット上に存在するが、その中には撮影された日付のデータが存在しないものも多く存在する。それらの画像はスキャンされたアナログ写真であったり、設定を誤ったデジタルカメラで撮影されたものである。撮影された日付が判明すればデータセットの質の向上や情報検索の多くの恩恵もたらされる。本研究ではそのような撮影日が不明な写真の撮影日の自動推定のための新しい手法を提案する。ニューラルネットワークのための順序一貫性を保った順序回帰手法を用いて、写真の撮影された年の推定を行った。また、入力としてオリジナル画像のみならずセマンティックセグメンテーションを用いたマスク画像を追加して精度の向上を図った。

キーワード 画像認識、順序回帰、セマンティックセグメンテーション

1 はじめに

現在、多くの画像がオンライン上でアクセス可能である。博物館や図書館などの機関は所蔵品をデジタル化しネット上で公開しており、例えばアメリカのスミソニアン博物館は4万点を超える所蔵品をデジタル化しWeb上で一般公開している。また、InstagramやFlickrのようなソーシャルネットワークサービスでは、そのサービス内でユーザーがアップロードしている。特にSNSの流行は画像のアップロード数の増加を急速に加速させており、Instagramの月間アクティブユーザーは10億人超、Facebookの月間アクティブユーザーは約25億人（2019年第4半期）¹と発表されておりSNSの勢いと同時に画像のアップロード数が凄まじいものであることが容易に想像される。

博物館や図書館などが公開している画像であれば、その画像に関する詳細な情報（誰がどのような目的でいつ作成したのかなど）が画像と共に明記されている。また、SNSにアップロードされる画像の多くも撮影された時点でデジタル画像であるために画像のメタ情報（撮影した端末、撮影日、撮影場所など）が付与されている。しかしながら、撮影日などのメタ情報が欠損している画像も中には存在している。例えば、アナログ写真をスキャンした画像や端末の設定を間違えたデジタルカメラの写真などは撮影者・撮影日・撮影場所などの情報が欠損している。

本研究ではこのようなメタ情報の欠損した画像の撮影された年の推定に焦点を当てる。撮影年推定を可能にすることで受けられる恩恵が大きく2つ存在する。

1つ目はデータのまとまりとしての質の向上である。その一例として、撮影された年が不明なデータが混ざったデータセットの存在を考えてみる。撮影年推定が不可能であれば、写っているオブジェクトごとでカテゴリ化する程度の事が可能である。しかし、撮影年の推定が可能であれば撮影年でソートする

ことや「1930年代のニューヨーク」といったクエリをリクエストすることも可能となり、情報検索の分野にも貢献できる。つまり、データとしての運用のしやすさや、データとしての価値を向上させることができる。

2つ目は歴史的に重要な写真の保護である。撮影した年が推定されることによって歴史的な価値が判明することもある。写真はその時代の文化を表す様々な情報をもたらす。一見すると何の変哲も無い写真であったとしてもその写真に写っている建物や人、ファッションや食べ物などからその時代の文化や技術が読み取れる。例えば、一般的に1950年代に流行したことが知られているファッションをした人が写っている写真が1940年代に撮影されたことが判明した状況を考える。そのような格好をする文化が実はすでに1940年代には存在することがその写真から読み取れる。この場合、その写真は歴史的に重要な事を証明する一枚となり保存すべき対象となる。人類の歴史を保存する一助として撮影年推定は重要なツールとなる。

写真のアーカイブを作り、歴史的に重要な写真を保存するという取り組みの一環としてFlickr²はThe Commons³というプロジェクトを2008年からアメリカ議会図書館と共に推し進めている。The Commonsでは古い写真を世界中から集めて誰でもアクセスできる写真のコレクションを作っている。

これ以降の部分では関連研究について述べ、提案手法を説明した後、予備実験について書く。最後に考察を述べる。

2 関連研究

2.1 写真の撮影日推定

Palermo [1]らはニューラルネットワークが流行する以前にサポートベクターマシンを用いてアナログ写真画像の撮影年推定を行った。この研究ではデータセットを集め、色の分布などの画像自体の特徴や時代ごとのカメラの特徴などを特徴量として取り出し、写真の年代判定をクラス分類の問題として解

1 : <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

2 : <https://flickr.com/>

3 : <https://www.flickr.com/commons>

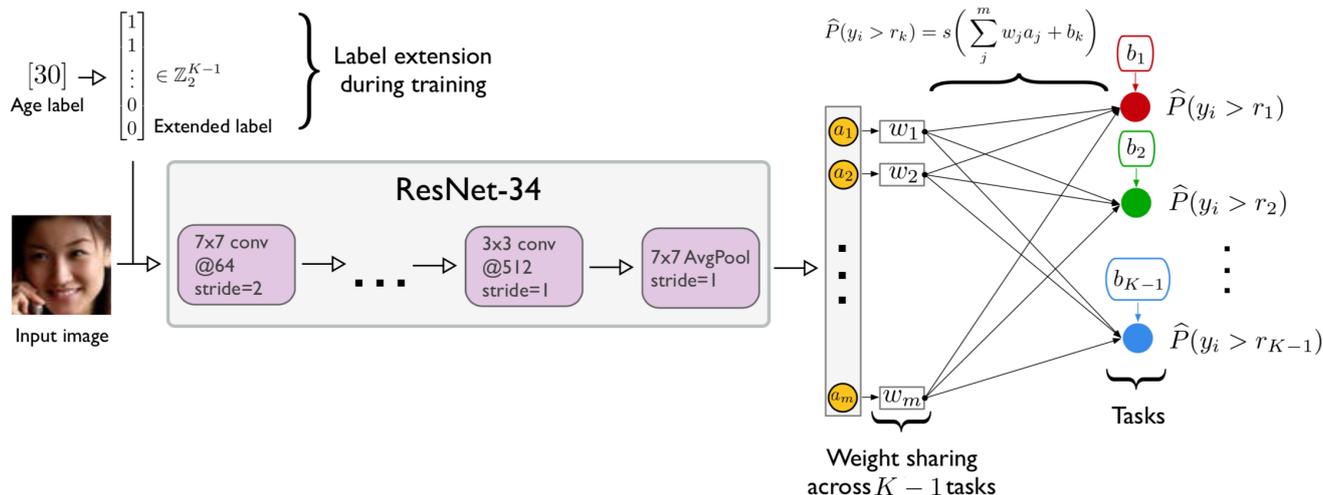


図1 一貫性を保った順序回帰 [5]

いた。さらにその特徴量を活かして比較的新しい写真を古く見えるように加工したり、古い写真を新しく見えるように加工したりするというアプリケーションの作成までを行った。また Fernandor [4] らは Palermo の特徴量に加えて、撮影時の状況の光や影の具合を考慮した色の特徴量を入力として精度の向上を図った。Müller [2] らは Flickr から大量のデータを入手して新たなデータセットを作成し GoogleNet という画像認識のモデルを用いて多クラス分類、そして回帰問題の両方でこの撮影年推定の問題に取り組んだ。Palermo との比較はしていなかったが、人間より良い精度で推定を行う事ができていた。また、Martin [3] らは撮影年の推定を順序分類のアプローチで行った。K ランクの順序分類問題に対して、ある年より前に撮影された写真かその年以降に撮影されたものかを判定する分類器を K 個作成し、各分類器のスコアから最も確率の高い年を推定値とする手法を提案した。

2.2 画像の検索

はじめにの部分でも触れたように画像の検索の分野においても様々な研究が見られる。Dias G [6] らは画像検索のクエリに暗に含まれる時間の情報 (“第二次世界大戦” など) をうまく抽出し、画像検索の精度の向上を図った。また、Chatzichristofis [7] らは Fuzzy Color and Texture Histogram と呼ばれる特徴量を抽出することで歪みやノイズが含まれた画像の場合でも適切な検索を可能にした。

3 提案手法

3.1 一貫性を保った順序回帰

本研究では、撮影年の推定のために順序回帰手法を用いた機械学習を CNN を用いて行う。順序回帰とは正解ラベルに順序が存在するクラス分類で順序分類とも呼ばれる。順序回帰は様々な場面で活用されている。人の年齢推定 [9] や病気の進行の程度 (アルツハイマーやクローン病など) の推定、最近では広告や推薦の技術としても用いられている。

撮影年の推定は順序回帰ではなく多クラス分類や通常の回帰の問題として扱えそうである。しかし、多クラス分類の場合には各クラスは独立しているはずだが、撮影年には順序関係が存在しているため多クラス分類の問題には向いていない。また通常の回帰の場合には、撮影年のラベルを数値として扱う。写真の特徴は撮影年に依存して変化していくが、それは定常的なものではないので過学習を引き起こしてしまいやすい。例えば、街の風景の写真について考えてみる。街の風景の写真は時代が現在に近づけば近づくほど画質も良くなるし近代的な建物も増えていくが、時代によっては戦争などで荒んだ風景となることもありえる。このように時代の進行とともに定常的に街の風景が変化していく訳ではなく様々な要因を伴って変化していくため、回帰問題として解き辛くなっている。このような理由から本研究では順序回帰を用いて撮影年の推定を行う。

従来の順序回帰手法を用いた機械学習では順序の矛盾が生じるという問題がある。K ランクの順序回帰問題として人間の顔画像から年齢を推定するタスクを例として考えてみる。K ランクの順序回帰問題は K-1 個の 2 値分類問題として解かれる。各分類機 $i (i \in 0, 1, \dots, K-1)$ は入力されたデータがランク i より大きいかどうかを判断する。制約が何も無い場合、 i 番目の分類器が 30 歳より歳を取っていると推定しているにも関わらず、それより前の判定機が 20 歳よりは若いと判定してしまうことがある。これが順序の矛盾である。

そこで順序の矛盾を生じさせないために、ニューラルネットワークのための順序一貫性を保った順序回帰手法を用いる。この手法では CNN の出力に共通の重みを掛け合わせ、それぞれ独立したバイアスを加えることで順序の一貫性を保ったまま順序回帰による推定を行うことができる [5]。

3.2 ResNet

ニューラルネットワークのモデルとして ResNet34 を用いた。ResNet は 2015 年に Microsoft Research が考案したモデルである。現在の画像関連の学習モデルの多くが ResNet の亜種で

あるので、実験には様々なモデルのベースとなる ResNet を用いることとした。

3.3 セマンティックセグメンテーション



図2 マスク画像

また入力する特徴の1つとしてオリジナルの画像のみではなくセマンティックセグメンテーション [8] を用いたマスクを追加する。セマンティックセグメンテーションは、画像系ディープラーニングの一種で、画素レベルでそれが何かを認識するタスクを行う手法である。図2はマスク画像の例である。



図3 戦時中1



図4 戦時中2



図5 群衆



図6 子供たち

カメラで撮影されるものは時代によって変わるものである。図3や図4で見られるように、戦時中であれば崩壊した建物の写真が多く見られるし、また古い写真に多く見られるのは図5のような多くの人が集合しているものである。それに比べて新しい写真になればなるほどカメラが一般大衆に広まっていくので、図6のように人物をアップで撮影するような写真が多くなる。これらの画像からわかるように、写真の構図というのはカメラの普及やその時代の情勢により大きく変化する。そのような特徴を学習の際に考慮するためにセマンティックセグメンテーションを入力に加えることにした。

4 予備実験

先ほど説明を行った一貫性を保った順序回帰手法が写真の撮影年推定のタスクにおいてどれほどの精度が見られるのかを確認するために実験を行った。学習モデルは ResNet34 で出力の

1つ前の層の重みを共有することで順序の一貫性を保持する順序回帰手法を適用した。

4.1 データセット



図7 Date Estimation in the Wild Dataset ⁴ [2]

データセットは Date Estimation in the Wild Dataset を用いた。データセットに含まれる画像の例を図7に示す。このデータセットは1930年から1999年までの画像を Flickr からダウンロードしてきたもので、画像は計1029710枚となっている。その内、各年16枚ずつ計1120枚がテストデータである。また、白黒画像とカラー画像での学習の差を確認するための簡易的なデータセットとして各年1000枚ずつをランダムに取得してきた70000枚の画像も作成した。

4.2 結果

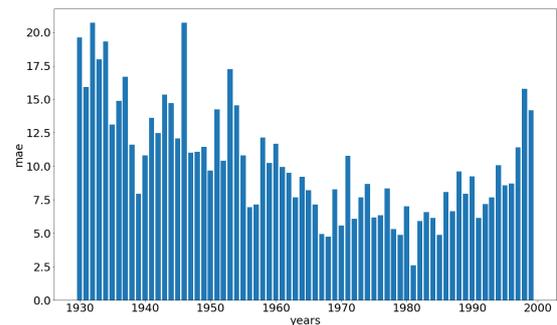


図8 各年の MAE

表1は26歳から58歳までの7人のアノテーターの評価、提案手法で学習したモデル、そして GoogleNet で学習を行ったモデル [2] それぞれの Mean Absolute Error を示している。学習はどちらもテストデータを除いた約100万枚の画像で行った。表2は70000枚のデータセットを用いてグレースケールで学習したモデルとカラー画像のまま学習したものを比較評価した表である。

また図8は提案手法で学習したモデルをテストデータで評価した際の各年の MAE をグラフで表現したものである。

5 考察

5.1 モデルの評価

本実験ではランダムに初期化された重みを使用したが、事前学習済みのモデルを使用すればより良い評価が得られることが

	Human	Proposed Method with 1M images	GoogleNet with 1M images
MAE	10.9	10.41	7.5

表 1 予備実験の結果

	Proposed Method with 70K black and white images	Proposed Method with 70K color images
MAE	15.74	11.41

表 2 モノクロ画像とカラー画像の比較

見込まれる。また、インプットとなる画像を 120x120 の画像にリサイズして学習を行ったが、もう少し大きなサイズで学習すればより多くの情報で学習を行うことができるので、精度は向上するであろう。

5.2 データセット

5.2.1 モノクロ画像

データセットには白黒画像とカラー画像の両方が混ざっている。モノクロ画像で学習するときとカラー画像で学習するときでは精度にどれほどの差が生まれるのかを確認するために、全ての画像を白黒画像にして学習したものと白黒画像とカラー画像が混ざったままで学習したものを比較した。結果として白黒画像のみで学習を行った際には精度が落ちることが確認できた。テストデータの白黒画像とカラー画像の割合は約半分であり、また時代が古いほど白黒画像の比率は高くなる。これらを考慮すると古い写真の判定の精度が低くなってしまうので、何らかの対策を考える必要があると考えられる。

5.2.2 封筒・手紙

データセットの中には手紙や封筒のスキャン画像が入っている。テストセットには手紙や封筒のスキャン画像は含まれていないため、訓練データから取り除くべきである。手紙や封筒のようなデータを除くためには、Flickr のタグから”envelope”や”letter”が含まれているような画像を取り除くことやセグメンテーションのマスク画像から判断することなどが考えられる。

5.2.3 データオーグメンテーション

学習に際してデータオーグメンテーションなどを行っておらず、過学習の状態に落ちるのが比較的早かった。まだまだ、データセットに関しても改善の余地が見られる。

5.3 セマンティックセグメンテーション

提案手法ではセマンティックセグメンテーションに触れたが今回の予備実験ではまだ導入することができていないので確認する必要がある。

文 献

- [1] Palermo F., Hays J., Efros A.A. (2012) Dating Historical Color Images. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577. Springer, Berlin, Heidelberg
- [2] Müller E., Springstein M., Ewerth R. (2017) “When Was This Picture Taken?” – Image Date Estimation in the Wild. In: Jose J. et al. (eds) Advances in Information Retrieval.

- ECIR 2017. Lecture Notes in Computer Science, vol 10193. Springer, Cham
- [3] Paul Martin, Antoine Doucet, and Frédéric Jurie. 2014. Dating Color Images with Ordinal Classification. In Proceedings of International Conference on Multimedia Retrieval (ICMR '14). Association for Computing Machinery, New York, NY, USA, 447–450. DOI:https://doi.org/10.1145/2578726.2578790
- [4] B. Fernando, D. Muselet, R. Khan and T. Tuytelaars, “Color features for dating historical color images,” 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 2589-2593.
- [5] Wenzhi Cao, Vahid Mirjalili, Sebastian Raschka, “Rank-consistent Ordinal Regression for Neural Networks,” 2019
- [6] Dias G., Moreno J.G., Jatowt A., Campos R. (2012) Temporal Web Image Retrieval. In: Calderón-Benavides L., González-Caro C., Chávez E., Ziviani N. (eds) String Processing and Information Retrieval. SPIRE 2012. Lecture Notes in Computer Science, vol 7608. Springer, Berlin, Heidelberg
- [7] S. A. Chatzichristofis and Y. S. Boutalis, “FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval,” 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services, Klagenfurt, 2008, pp. 191-196.
- [8] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, Antonio Torralba; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 633-641
- [9] Paul Martin, Antoine Doucet, and Frédéric Jurie. 2014. Dating Color Images with Ordinal Classification. In Proceedings of International Conference on Multimedia Retrieval (ICMR '14). Association for Computing Machinery, New York, NY, USA, 447–450. DOI:https://doi.org/10.1145/2578726.2578790