

手作り作品の紹介画像と説明文に基づく作家識別

田中 大生^{1,a)} 三宅 悠介^{1,3} 峯 恒憲^{1,2,b)}

概要: 本研究では、手作り作品の写真画像と紹介文のデータをもとに、作家識別を行うモデルを提案する。作品の作家識別性能の高いモデルは、その作品の特徴説明だけでなく、同じ特徴を持つ作品の検索や推薦にも利用できる。本研究では、手作り作品のデータのうち、写真画像、紹介文（説明文、タイトル）を用いた作家識別を行うモデルを提案する。評価実験の結果、写真画像もしくは紹介文を入力として扱うモデルと比較して、MicroF1 値で 2% および 5% の精度向上ができた。この実験から、手作り作品の画像データの情報を紹介文で補足することで作家識別の精度が向上する事が分かった。その後、写真画像の内容を紹介文で説明することを目標に、物体検出を行った結果を用いて作家識別を行った。評価実験の結果、物体検出の結果は、作家識別精度と関連が強いことは確認できたものの、物体検出の結果だけを用いて、作家識別を行うことは困難であることも確認した。

Identification of the creator based on the featured image and description of the handmade work

1. はじめに

Eコマースの普及により、手作り作品の市場も急速に拡大しつつある。一方、世に送り出される手作り作品の数が増えたことから、作家名やジャンルなどの詳細な情報を欠く作品も多くなっている。そのため、作品の外観から自動で詳細情報を判別できる事は、作家名の判別だけでなく、作品の作風などの特徴に加え、作家の芸術的スタイルの理解などに役に立つ。

作品の外観情報を利用する画像処理分野の研究では、Convolutional Neural Network (CNN) が作家識別、顕著性検出 (Saliency Detection)、視覚的記述生成などで高い適性を示している。一方、手作り作品では、同じ作品名でも外見に大きな差が出やすいことから、作品の外観情報だけを利用して、作品が表す内容や作家の名前などを推測す

ることは容易ではないと考える。

そこで本研究では、手作り作品の作家識別を高精度に行うことを目的に、作品の外観情報にあたる写真画像を利用するモデル、作品の紹介文（タイトルおよび説明文）データを利用するモデル、ならびに、2つのデータ双方を利用して作家識別を行うアンサンブルモデルの提案を行う。作品の作家識別性能の高いモデルは、その作品の特徴説明だけでなく、同じ特徴を持つ作品の検索や推薦にも利用できる。そのため、作品の外観情報だけでなく、紹介文データの有効性を作家識別というタスクを通して検証することは価値がある。さらにアンサンブルモデルを通して、2つのデータ双方の間の関連性について確認することは、同じ特徴を持つ作品の検索や推薦サービスを実現する上で有用である。

提案モデルの有効性を検証するため、実際の手作り作品サイトから収集したデータを利用して実験を行った。実験の結果、作品の写真画像および紹介文データを単体で作家識別に利用するモデルと比べ、アンサンブルモデルは、それぞれ F1 値で 2%、5% の精度向上を得た。これにより、写真画像および紹介文データは、相互に補完する情報を有することを確認した。

さらに手作り作品の写真画像の属性情報をより詳しく捉えるため、まず、写真画像から学習済みモデルで物体検出

¹ 九州大学大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University, Nishi, Fukuoka 819-0395, Japan

² 九州大学大学院システム情報科学府
Factory of Information Science and Electrical Engineering, Kyushu University, Nishi, Fukuoka 819-0395, Japan

³ ペパボ研究所
Pepabo R&D Institute, GMO Pepabo, Inc., Chuoku, Fukuoka 810-0001, Japan

a) tanaka.taisei.129@s.kyushu-u.ac.jp

b) mine@ait.kyushu-u.ac.jp

を行い、検出された物体ラベル名を利用して、作家識別のエントロピーを計算した。ついで、写真画像だけを利用した作家識別モデルの精度と比較したところ、エントロピーの大きさとモデルの精度との間には、負の相関（すなわちエントロピーが小さいほど、識別精度は高い）があることを確認した。これにより、写真画像に基づく作家識別精度は、写真画像からの物体検出モデルが捉える特徴の精度と関連が強い可能性があることを確認できた。一方、物体検出により獲得したラベルだけで作家識別モデルを構築した場合には、作家識別精度が大きく下がることから、物体検出時に作家識別に必要な多くの情報が失われている可能性が高いこともわかった。

以下、2節では関連研究について述べ、本研究の立場を明らかにする。3節では、提案手法と、実験で利用するデータセットについて述べる。4節では、提案手法の有効性を検証するための実験結果について考察する事に加え、物体検出のを利用した作家識別モデルで作品の特徴をより詳細に捉えることを試みた。最後に5節でまとめと今後の課題について述べる。

2. 関連研究

CNNを用いて画像から作家を識別する取り組みは数多く存在する。たとえばJangtjik[1]らは入力画像の解像度別が違う3つのモデルを組み合わせる新しい画家同定の手法を提案した。この研究では、13作家の1300枚の作品画像をデータセットとして利用し、マルチスケールピラミッドを構築する。第1層は固定サイズのCNN、第2層と第3層の縮尺画像は第1層の縮尺画像の4倍と16倍の解像度を持つ。この手法は上位2位までの検索において88.08%のrecallを達成している。

Guilherme Folego[2]らはVincent van Goghの絵画を識別するために、画像から画家の視覚的パターンを直接抽出するCNNと、最終判定に融合法を用いた分類器を提案した。各絵画をバッチ分割して個別に分類した後、最終的な識別結果を求めた結果、投票方式を上回る結果を得た。

深層学習を用いた自然言語処理による分類タスクを行う研究では、例えば、Ali Alessaら[3]は、FastTextを用いてSNS上の投稿をインフルエンザ関連の投稿と非関連の投稿に分類する手法を提案した。提案手法の有効性を検証した結果、SNSの投稿のような非構造化データを使用するインフルエンザ疾患監視システムの精度と効率の向上に成功している。

M. Cliche[4]はSNS上の書き込みのセンチメント分類を行う手法として、2つの深層学習技術であるCNNとLSTM[5]のアンサンブル学習を提案した。この研究ではラベル無しデータの一部で使ってDistant supervision[6]で埋め込みを微調整した後に、SemEval-2017 Twitterデータセットでモデルを学習している。この手法はTwitter感情

分析コンテストSemEval-2017の全項目で1位を達成している。

画像と文章を組み合わせた入力データで分類タスクを行う研究では、例えば、Alec Radfordら[7]は、インターネットから収集した4億の画像と文章のペアをデータセットとして、画像に付ける注釈を予測する実験を行った。その結果、このモデルは多くのタスクで自明であり、データセット固有の訓練を必要とせずに完全に教師あり学習と同等の性能であることを示した。

本稿では、944作家、109,685の作品からなる手作り作品のデータセットを構築、利用し、作品に付随する写真画像および紹介文の双方を入力として作家識別を行うモデルを提案する。

3. 提案手法と実験内容

3.1 実験で使用するデータセット

本稿の実験で使用するデータセットは、ハンドメイドマーケットminne (<https://minne.com/>) から収集した写真画像と紹介文データで、下記の条件を満たす物である。

- (1) 2012年4月3日11:03から2019年9月1日23:31の期間に出品された作品
- (2) カテゴリがコサージュ・ブローチ
- (3) 同一作家を除去
- (4) 作品数50未満の作家を除去

対象作家の作品数の最大値は4,367、最小値は50で、標準偏差は177.7である。

3.2 評価尺度

提案手法の評価には、Overall Accuracy (以下、Accuracy) (式1参照) ならびにMicroF1値 (式2参照)、MacroF1値 (式3参照) を利用する。

$$Accuracy = \frac{T}{N} \quad (1)$$

$$MicroF1 = \frac{2PR * PE}{PR + PE} \quad (2)$$

$$MacroF1 = \sum_{i=1}^n \frac{2PR(i) * PE(i)}{PR(i) + PE(i)} \quad (3)$$

N = 全作品数

n = 全作家数

T = 予測作家が実際の作家と一致する作品数

PR = 作家識別の適合率

RE = 作家識別の再現率

$PR(i)$ = 作家 i の作品の適合率

$RE(i)$ = 作家 i の作品の再現率

3.3 作品画像を対象とした作家識別モデル

CNNモデルの一つであるVGG16[8]を利用し、作品の

写真画像を入力とする作家識別モデルを構築する。VGG16は、特徴量抽出を行う畳み込み層13層と全結合層3層の計16層からなる。提案モデルでは、畳み込み層の初期値として、転移学習によりImageNet (<https://image-net.org/>)で学習した重みを利用する。全結合層(Dense層)は1層とし、出力層の各次元には異なる作家を対応付け、Softmaxにより、作家の識別確率を出力する。そのため、出力層の次元数は使用するデータセットの作家数944とする。モデルの構造を図1に示す。なお、後述するモデルも、同じ各次元と作家との対応関係を持つ出力層を用いる。

学習/検証用と評価用データは、各作家の作品画像を4:1の比率で分割して、5分割交差検証で評価する。このとき、学習/検証用データは更に5等分し、学習用と検証用として4:1の比率で分割し、学習用と検証用データを交差検証の要領で入れ替えた5通りの学習モデルを構築し、それぞれの学習モデルを利用した評価結果を求め、その平均を識別結果とする。

モデルの学習/検証にはKerasを、パラメータチューニングにはOptunaを、そして最適化アルゴリズムには確率的勾配降下法を用いた。エポック数は100で、バッチサイズは256に設定した。ハイパーパラメータのチューニングにより、モデルの全結合層は4096次元の1層、learning_rateは0.005、momentumは0.97とした。

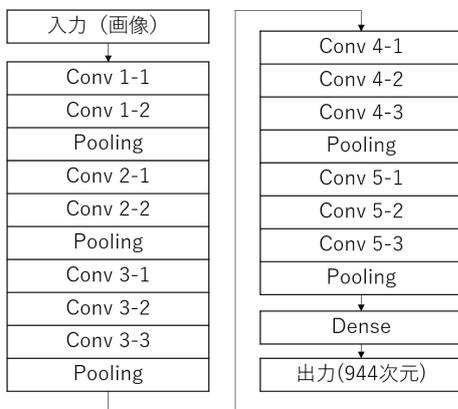


図1 画像を用いた作家識別の単独モデル

3.4 紹介文を対象とした作家識別モデル

紹介文データ(タイトルと説明文)は、MeCab (<https://taku910.github.io/mecab/>)を用いて分かち書きしたものをFastText[9]の入力として作家識別モデルを構築する。ハイパーパラメータはdim=100, epoch=1000とし、他はデフォルト値である。モデルの構造を図2に示す。

紹介文を用いた作家識別モデルを構築する際に、FastTextを選ぶ際には、事前に以下の4つの方法で行った。ただし、実験で用いるデータセットとは別に学習用・検証用・評価用として3:1:1の比率で分割したものを学習モデ

ルの構築と評価に用いた。

- (1) 紹介文をWord2Vecで100次元ベクトル化し、LightGBM[11]で作家識別。
- (2) 紹介文をDoc2Vecで300次元ベクトル化し、LightGBMで作家識別。
- (3) 全紹介文中の出現回数上位5000単語で、各紹介文をBoWベクトル化し、各次元の値を対応する語のTF-IDF値としてLightGBMで作家識別。
- (4) FastTextで作家識別

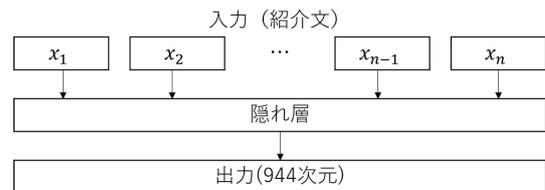


図2 作品の紹介文を用いた作家識別モデル

LightGBMのパラメータはlearning_rateを0.1, num_leavesを23, min_data_in_leafを1, num_iterationを100とした。手法1から手法4のAccuracyが、それぞれ0.07050, 0.5005, 0.8364, 0.8806であったことから、識別精度のもっとも高い、手法4のFastTextを採用した。

3.5 画像と紹介文の双方のデータを用いた複合学習モデル

作品の写真画像と紹介文それぞれの単独モデルを使った作家識別の出力を結合し、作家識別を行う。モデルの構造を図3に示す。写真画像と紹介文のそれぞれに対して5通りの学習/検証データで学習を行った合計10のテストデータの出力の加重平均を出力とした作家識別を行う。加重平均の比率は、画像単独モデルと紹介文単独モデルを1:9, 2:8, ..., 9:1とした場合を互いに比較し、評価値が最大の比率を提案手法として採用する。画像と紹介文の単独モデルの出力は各要素を対応する作家である確率とした944次元配列であり、複合モデルの出力もまた、単独モデルの出力を平均した944次元配列である。

3.6 物体検出の出力を用いた学習モデル

物体検出モデルであるSSD(Single Shot multibox Detection)[10]の学習済みモデルを利用する。SSDは写真画像内の物体を矩形で囲み、矩形毎に物体検出とクラス分類を平行して行い、物体ラベルとその信頼率(以下確率)を求める。SSDによって行われた物体検出の例を図4に示す。学習済みモデルは約33万枚の大規模なカラー画像のデータセットであるCOCO Dataset (<https://atmarkit.itmedia.co.jp/ait/articles/2109/08/news026.html>)で学習されたものを使用し、図5に示す学習済みモデルが識別可能な92個の物体ラベルを検出する。

各次元に異なる物体ラベルを対応付けた92次元ベクト

ルにより、各作品を表現する。ベクトルの次元の値は、その次元に対応するラベルの信頼確率とする。同一のラベルをもつ複数の物体が検出された場合、その最大の信頼確率を、その次元の値とする。なお物体検出時には、確率 0.2 以上のものだけを検出し、0.2 に満たないラベルの次元の値は 0 とする。作品の特徴ベクトルを入力として LightGBM[11] で作家識別を行う。パラメータは learning_rate を 0.1, num_leaves を 23, min_data_in_leaf を 1, num_iteration を 100 とした。

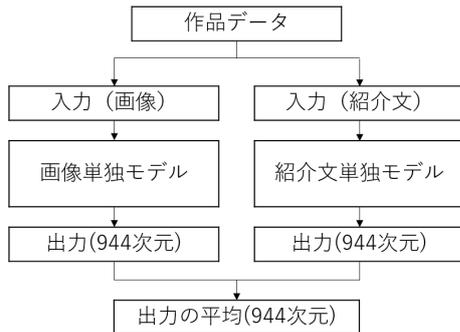


図 3 作品の紹介文を用いた作家識別モデル



図 4 物体検出の例

3.7 物体ラベルの組み合わせエントロピーを用いて選択した作品での評価値検証

特定の作品のみの場合で、画像単独モデル、紹介文単独モデル、画像と紹介文の複合学習モデル、物体検出モデルの作家識別性能の変化を調べる。エントロピーを導入する目的は、物体ラベルによる作家識別がしやすい作品を発見することである。各作品の作家と SSD で検出された物体ラベルの組み合わせエントロピー H (式 4 参照) を計算する。各学習モデルの全作品での識別結果から閾値未満のエントロピーを持つラベルの組み合わせが検出された作品を取り出し、評価値を調べる。閾値は [1,2,3,4,5] で実験を行った、閾値ごとの対象作品の作家別作品数の統計量を表 1 に示す。参照この実験によって検出された物体ラベルから作家識別しやすい作品を発見できるかどうかを検証する。

None, 人間, 自転車, 車, バイク, 飛行機, バス, 電車, トラック, ポート, 信号, 消火栓, 標識, 停止, メーター, ベンチ, 鳥, 猫, 犬, 馬, 羊, 牛, 象, くま, シマウマ, キリン, 帽子, リュック, 傘, 靴, めがね, バッグ, ネクタイ, スーツケース, フリスビー, スキー, スノーボード, スポーツ, ボール, カイト, バット, グローブ, スケボー, サーフィン, テニス, ラケット, ボトル, プレート, グラス, カップ, フォーク, ナイフ, スプーン, ボウル, バナナ, リンゴ, サンドイッチ, オレンジ, ブロッコリー, ニンジン, ホットドッグ, ピザ, ドーナツ, ケーキ, 椅子, ソファ, 鉢植え, ベッド, 鏡, テーブル, 窓, 机, トイレ, ドア, テレビ, パソコン, マウス, リモコン, キーボード, 携帯, レンジ, トースター, シンク, 冷蔵庫, ブレンダー, 本, 時計, 花瓶, ハサミ, テディベア, ドライヤー, 歯ブラシ

図 5 使用した SSD の学習済みモデルが識別可能な物体ラベル一覧

表 1 閾値ごとの対象作品の作家別作品数

閾値	全作品数	最大値	最小値	標準偏差
無し	109685	4367	50	177.7
5	54482	2952	8	108.8
4	47053	2868	5	102.0
3	38304	2364	2	83.03
2	19420	2059	1	70.55
1	6011	1237	1	54.65

$$H(LC) = - \sum_{i=1}^n P(LC, i) \log P(LC, i) \quad (4)$$

$H(LC)$ = 物体ラベルの組み合わせ LC のエントロピー

n = 全作家数

LC = 物体ラベルの組み合わせ

$P(LC, i)$ = 物体ラベルの組み合わせ LC の作品に含まれる作家 i の作品の割合

4. 実験結果

画像と紹介文の複合モデルにおいて最良の加重平均の比率を探すために、画像モデルと紹介文モデルの比率を 1:9, 2:8, ..., 9:1 の場合において評価値を比較した。この実験結果を表 2 に示す。全評価値が最大の比率 5:5 を提案手法として採用する。

提案手法の有効性を調べるために、画像単独モデル、紹介文単独モデル、画像と紹介文の複合モデルそれぞれの評価値をデータセットの全作品において比較した。この実験結果を表 3 に示す。画像と紹介文を複合的に用いる提案手法が、それぞれ単独で用いる場合より高い評価値を示しており、組み合わせ利用することの効果を確認できた。提案手法の有効性をさらに分析するために、単独モデルと複合モデルで正誤が変化した作品数についての分析結果を表 4 に示す。単独モデル両方で不正解かつ複合モデルで正解の作品数は 99, 単独モデル両方で正解かつ複合モデルで不正

表 2 単独モデルの比率を変化させた場合の画像と紹介文の複合モデルの実験結果

比率 (画像 : 紹介文)	Accuracy	MicroF1 値	MacroF1 値
1 : 9	0.8115	0.8115	0.7586
2 : 8	0.8387	0.8387	0.7923
3 : 7	0.8757	0.8757	0.8380
4 : 6	0.9077	0.9077	0.8768
5 : 5	0.9299	0.9299	0.9016
6 : 4	0.9214	0.9214	0.8891
7 : 3	0.9072	0.9072	0.8691
8 : 2	0.8940	0.8940	0.8516
9 : 1	0.8844	0.8844	0.8397

表 3 全作品での比較実験結果

実験	Accuracy	MicroF1 値	MacroF1 値
画像単独モデル	0.7959	0.7959	0.7400
紹介文単独モデル	0.8793	0.8793	0.8338
提案手法	0.9299	0.9299	0.9016
物体検出モデル	0.09056	0.09056	0.05893

表 4 単独モデルと複合モデルの正誤変化作品数

	画像→複合	紹介文→複合
T → T	86458	95563
F → F	6818	6838
F → T	15504	6399
T → F	905	885

解の作品数は 0 だった。提案手法を用いることで全体の評価値は向上したが、単独モデルで正解かつ提案手法で不正解の作品が存在した。これはアンサンブル学習の手法が最適化されていないことが原因と考えられるため、出力を平均するよりも適した手法を探すことが今後の課題である。

画像と紹介文の複合モデルから、さらに写真画像に映っている物体を詳細に理解するために、物体検出を利用したモデルで同様に実験を行い、評価値を前述の 3 つのモデルと比較した。この結果も表 3 に示す。表 3 からわかるように、物体検出の結果を利用したモデルは全ての評価値が他のモデル比べて大幅に低かったことから、この手法での作家識別が困難であることが分かった。

これらの結果を受けて、物体検出を用いた学習モデルの利用可能性を探るために物体ラベルの組み合わせエントロピーが閾値 [1,2,3,4,5] 未満の作品を対象に同様の実験を行った。実験結果を表 5 から 8 に示す。

表 5 の結果から、物体ラベルの組み合わせエントロピーが小さい作品になるにつれて画像単独モデルの Accuracy と MicroF1 値が向上することが確認できたことから、作品の内容の違いが作家識別精度に大きく関係する可能性が強いことを想定される。また、閾値 1 の Accuracy と MicroF1 の伸びが大きかったことから、画像単独モデルは特定の物体ラベルの組み合わせにおいて高い性能を有していると考えられる。

表 5 物体ラベルの組み合わせエントロピーが閾値以下の作品の画像単独モデルでの実験結果

閾値	Accuracy	MicroF1 値	MacroF1 値
5	0.8182	0.8182	0.7502
4	0.8223	0.8223	0.7484
3	0.8280	0.8280	0.7475
2	0.8467	0.8467	0.7522
1	0.9590	0.9590	0.7950

表 6 物体ラベルの組み合わせエントロピーが閾値以下の作品の紹介文単独モデルでの実験結果

閾値	Accuracy	MicroF1 値	MacroF1 値
5	0.8939	0.8939	0.8391
4	0.8950	0.8950	0.8372
3	0.8967	0.8967	0.8358
2	0.9055	0.9055	0.8337
1	0.9489	0.9489	0.7668

表 7 物体ラベルの組み合わせエントロピーが閾値以下の作品の画像と紹介文の複合学習モデルでの実験結果

閾値	Accuracy	MicroF1 値	MacroF1 値
5	0.9397	0.9397	0.9058
4	0.9405	0.9405	0.9051
3	0.9421	0.9421	0.9042
2	0.9499	0.9499	0.9050
1	0.9860	0.9860	0.9109

紹介文単独モデルでは表 6 に示すように、エントロピーの閾値が小さくなるにつれて Accuracy と MicroF1 値は増加しているが、MacroF1 値は逆に減少している。これは作品内容が変化する過程で作家間の正解率の差が開いてしまったことが原因と考えられ、紹介文単独モデルは必ずしも特定の物体ラベルの組み合わせにおいて高い性能を有しているとは言えないと考えられる。

作品の画像と紹介文の複合学習モデルでは、表 7 に示すように、評価値の変動が画像単独モデルと同じくエントロピーの閾値が小さい作品になるにつれて Accuracy と MicroF1 値が向上しており、作品の内容が変化しても単独モデルと比較して高い評価値を維持していることから、特定の物体ラベルの組み合わせにおいて画像と紹介文の単独モデルと比較しても高い性能を有していると考えられる。

物体検出モデルでは表 8 に示すように、エントロピーの閾値が小さい作品になるにつれて全評価値が上昇したが、依然として他の学習モデルを大幅に下回った。この結果から、限られた作品においても物体検出を用いた学習モデルでの作家識別が困難であることが分かった。

5. おわりに

5.1 まとめと今後の課題

本論文では、手作り作品の写真画像と紹介文を入力とした作家識別手法を提案した。945 作家の 109,685 の作品をデータセットとして実験を行った。その結果、画像や紹介

表 8 物体ラベルの組み合わせエントロピーが閾値以下の作品の物体検出モデルでの実験結果

閾値	Accuracy	MicroF1 値	MacroF1 値
5	0.1375	0.1375	0.07304
4	0.1519	0.1519	0.07597
3	0.1670	0.1670	0.08082
2	0.2092	0.2092	0.08897
1	0.4545	0.4545	0.1360

文を単独で用いた作家識別モデルと比べ、MicroF1 値で、2%および5%の性能向上を得られた。このことから、両方のデータを活用する重要性を確認した。また、手作り作品の属性をより詳しく捉えることを目標として、写真画像から物体検出を行い、検出された物体ラベルから作家を推定する実験を同様のデータセットで行った結果、物体検出モデルが捉える内容と、作家識別モデルの識別精度との間に、強い関連性が確認されたものの、この手法での作家識別は、残念ながら困難であることを確認した。

今後は、写真画像の特徴と、紹介文内の単語との関連づけを行うことで、写真画像データだけを用いて、提案手法の精度に近づけることを目指す。

謝辞 本研究の一部は科研費 JP21H00907, JP20H01728 の支援を受けた。

参考文献

- [1] K.A. Jangtjik, M. C. Yeh, and K.Hua.: *Artist-based Classification via Deep Learning with Multi-scale Weighted Pooling*, 24th ACM international conference on Multimedia (MM '16)., 635–639, 2016.
- [2] G. Folego, O. Gomes and A. Rocha.: *From impressionism to expressionism: Automatically identifying van Gogh's paintings*, 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 141-145, doi: 10.1109/ICIP.2016.7532335.
- [3] A. Alessa, M. Faezipour and Z. Alhassan.: *Text Classification of Flu-Related Tweets Using FastText with Sentiment and Keyword Features*, 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018, pp. 366-367, doi: 10.1109/ICHI.2018.00058.
- [4] Mathieu Cliche.: *BB_twttr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs*, ArXiv abs/1704.06125 (2017): n. pag.
- [5] Sepp Hochreiter and Jürgen Schmidhuber.: *Long Short-Term Memory*, Neural Comput. 9, 8 (November 15, 1997), 1735–1780.
- [6] Mintz, Mike and Bills, Steven and Snow, Rion and Jurafsky, Dan.: *Distant Supervision for Relation Extraction without Labeled Data* n Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2 (ACL '09). Association for Computational Linguistics, USA, 1003–1011.
- [7] Alec Radford and Jong Wook Kim and Chris Hallacy and Aditya Ramesh and Gabriel Goh and Sandhini Agarwal and Girish Sastry and Amanda Askell and Pamela Mishkin and Jack Clark and Gretchen Krueger and Ilya Sutskever.: *Learning Transferable Visual Models From Natural Language Supervision*, arXiv:2103.00020
- [8] Karen Simonyan and Andrew Zisserman.: *Very deep convolutional networks for large-scale image recognition*, Yoshua Bengio and Yann LeCun, editors, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.
- [9] Bojanowski, Piotr, et al.: *Enriching word vectors with subword information*, Transactions of the Association for Computational Linguistics 5 (2017): 135-146.
- [10] Liu, Wei and Anguelov, Dragomir and Erhan, Dumitru and Szegedy, Christian and Reed, Scott and Fu, Cheng-Yang and Berg, Alexander C.: *SSD: Single Shot Multi-Box Detector*, Lecture Notes in Computer Science Lecture Notes in Computer Science: 21–37, 2016.
- [11] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu.: *LightGBM: a highly efficient gradient boosting decision tree*, In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3149–3157.