筆跡画像を用いたうつ病重症度予測モデルの提案

田辺 寛人† 木村 昌臣‡

†‡芝浦工業大学大学院 〒125-8548 東京都江東区豊洲 3-7-5

E-mail: †ma23117@shibaura-it.ac.jp, ‡masaomi@shibaura-it.ac.jp

あらまし

本研究では，筆跡解析を通じてうつ病の重症度を判定するモデルを提案する．従来の診断方法は，医師と患者の主観的なコミュニケーションに依存しており,それが解釈の違いや診断までの時間の長期化につながる可能性がある．一方で，筆跡データを用いることで，客観的にうつ病を診断できると期待されている．本研究では，ResNet を用いて筆跡に向いたモデルになるようモデル構造を工夫した．また，Grad-CAM++を活用して筆跡画像内の注目領域を特定し，それらを正規化された筆跡速度と関連付けた．実験結果では，モデルの注目領域が速い速度よりも遅い速度の部分に集中していることが明らかになった．

キーワード SDGs，機械学習，医療・ヘルスケア，データマイニング，行動データ

1. はじめに

世界保健機関（WHO）によると，世界中で約 3 億人がうつ病を患っているとされている[1]．現在，うつ病の診断は患者と医師の会話を通じて収集された情報に基づいて行われている[2]．しかし，この診断方法には 2 つの問題がある．1 つ目は診断に時間がかかることであり，2 つ目は医師の主観的な判断に依存するため診断結果にばらつきが生じる可能性があることである．このように，医師と患者の対話に依存した診断には限界があるため，うつ病の重症度を客観的かつ効率的に判定するモデルが求められている．

これまでの研究では，うつ病の判定に音声[3][4]，表情[5][6][7][8]，行動[9][10]，脳波（EEG）[11][12][13]，および筆跡[14][15]などさまざまなデータが利用されてきた．しかし，これらの方法には多くの金銭的コストが伴う．例えば，音声の取得にはマイク，表情や行動の取得にはカメラ，脳波の取得には EGG ヘッドセット，筆跡の時系列データの取得にはタブレットや特殊なペンが必要である．また，これらの手法には情報取得のための設定や準備に時間的コストがかかるという課題もある．一方で，筆跡の画像を利用する方法は低コストであり，実施に場所を選ばない．そのため，筆跡画像を用いたうつ病判定の有効性を検証する必要がある．

筆跡からうつ病を推定することは容易ではない．理由は 2 つある．1 つ目は，筆跡科学が 1980 年代頃に疑似科学と見なされており，筆跡画像からうつ病の特徴を正確に抽出できるかが不明である点である．2 つ目は，データ量が少なく，データ作成に工夫を要する点である．

そこで本研究では，筆跡データを分割，オーギュメンテーションをすることでデータ不足を補った．また畳み込みニューラルネットワーク(CNN)の構造を筆跡に向いたモデルになるように工夫し，筆跡画像を用いてうつ病の重症度を予測するモデルを提案する．具体的には，スキップ接続を用いた ResNet [16]を採用し，筆跡画像からうつ病の重症度を推定するモデルを構築した．

本論文の構成は以下の通りである．第 2 章では本研究の貢献について述べる．第 3 章では関連研究を説明する．第 4 章では提案手法を詳細に記述し，第 5 章で実験結果を示す．第 6 章では考察を行い，第 7 章で本研究のまとめと今後の展望について述べる．

2. 先行研究

今まで，音声，表情，姿勢，脳波（EEG），および筆跡などの非言語的行動を用いたうつ病の判定が検討されてきた．音声を用いた手法では，臨床群と非臨床群の間で F0（基本周波数）やその変動性といった音響バイオマーカーが異なることが分析された．しかし，音声信号は外部のノイズの影響を受けやすく，特徴の信頼性が低下する可能性がある[3]．また，Williamson ら[4]による他の音声関連の研究では，音声における運動協調性をうつ病の重症度推定の指標として検討した．

表情分析や行動パターンの分析アプローチも有望な結果を示している．例えば，AVEC チャレンジでは，表情の微細な変化や音声を用いたモデルが，うつ病検出において 80%以上の精度を達成した[5][6]．さらに，Zhou らおよび Huang らの研究では，2D および 3D の顔データの高度な解析により，空間的および時空間的な顔の動態を同時に分析することで精度が向上した[7][8]．

姿勢や歩行パターンの分析も，うつ病の兆候との関連が報告されている．例えば，Michalak ら[9]は歩行パターンとうつ病の関係を見出し，Canales ら[10]は悪い姿勢がうつ病の再発と関連していることを示した．

脳波(EEG)を用いたアプローチも広く研究されている．Seal ら[11]および Spyrou ら[12]の研究では，うつ病患者の脳波パターンが健常者と有意に異なることが示され，感情状態の客観的な評価に有望であるとされている．しかし，EEG の測定には専門的な機器が必要であり，日常的な診断での利用は限られる．

筆跡に関する先行研究として 2 つの研究を説明する．Laurence ら[14]は，感情状態と筆跡を関連付けた最初の公開データベース（EMOTHAW）を作成した．彼らはランダムフォレストを用いて，129 人の参加者（女性 71 人，男性 58 人）が行った 7 つのタスク（5 つの描画，2 つの筆記）のデータを分析した．この研究では，うつ病患者は健常者に比べてタスクの完了に時間がかかることが明らかになった．Juan ら[15]も EMOTHAW データセットを用いて，時系列特徴に基づいてうつ病を予測した．タスクの総時間，ペンの浮遊時間，ペンの接地時間，ストローク数，および平均筆圧という 6 つの特徴を分析し，80.31%の精度でうつ病を検出した．これらの分析を行う際には，時系列特徴をフーリエ変換することで，時間的な詳細情報を取り入れ，予測精度を向上させた．上記の研究は筆跡に基づくうつ病予測の成功を示しているが，時系列データはデータ量が膨大になる傾向がある．

そのため，筆跡の画像を使用することで，機材コストと患者負担を最小限に抑えた手法が必要である．

3. 提案手法

3.1 前提

本研究では，Resnet を用い，入力として筆跡画像をモデルに入れ，出力としてうつ病の重症度スコアを 0 から 25 で推測するモデルを提案する．まず，モデルに入力する画像のデータセットの作成方法について説明する．本研究では，先行研究[14]のデータを使用する．このデータでは，4 つの単語を書く際に，x，y，タイムスタンプ，ペンの状態，方位，高度，筆圧の 7 項目が 1 つのファイルに記録されている．画像作成時には x,y 座標を使用し，出力する画像は下記になる．

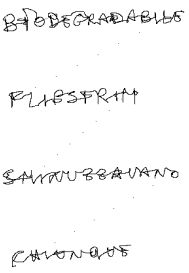

図 1　x 座標 y 座標の情報をもとにした画像

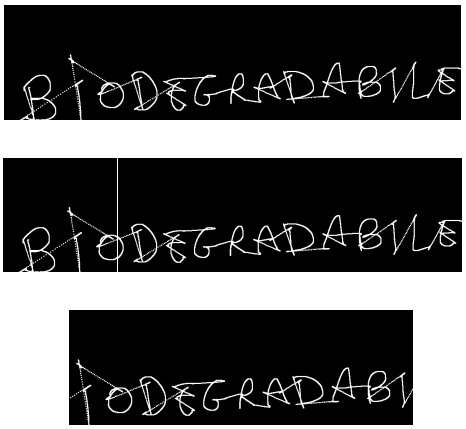

図 2　白黒反転画像と分割例

当初，4 つの単語の時系列データは 1 つのファイルに含まれていたため，このデータを単語ごとに個別のファイルに分離する必要があった．そのため，次の単語を書く時に，大きく左と下に動く行動が行われることに着目した.この移動を検知し,ファイルを分離した．その後，各単語はデータに基づいて 300x1200 の画像に変換した．手書き文字は黒を背景に白で描画した．そして，各画像は 4 つの 300x300 画像に分割した．理由は 2 つある．一つ目はデータ数が少ないため，単語を分割することでデータ数を増やすためである．2 つ目は，単語全体の筆記傾向（例えば，文章が上向きに傾いている）ではなく，対象者固有の文字の筆記傾向を得るためである．さらに，最初の分割で失われた可能性のある細部を捉えるために 3 つの中間画像を生成し，1 単語あたり合計 7 枚の画像を得た．

そして 0 から 25 までのすべてのうつ病重症度スコアにおいてデータ数を均等になるようにデータオーギュメンテーションを行った．こうして，モデルの入力として 300x300 の画像からなるデータセットを作成した．

3.2 モデル

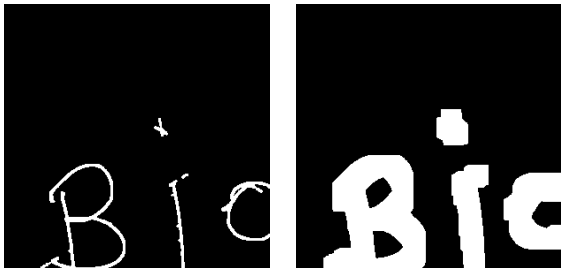

図 3 オリジナル画像と太線化した画像

我々のモデルは ResNet 層と全結合層(FC 層)の組み合わせで構成されている.具体的には 5 つの ResNet 層と 5 つの FC 層を使用している．CNN は通常，手書き解析のような画像処理タスクに効果的であるため，当

初は CNN を使用した．しかし，CNN モデルを利用した時に，うつ病の重症度予測値が一定になるという問題が起こった．この問題は，個々の手書き特徴を一般化し，本質的な情報を除去する Max pooling 層に起因していた（図 3）．今回採用したデータセットは，129 人の参加者が同じ 4 つの単語を書いていたため，Max pooling を行うと，各個人の書き癖などの情報が欠落し，同じ特徴を得てしまっていた．Max pooling は，上記の画像のように，太線化の効果をもたらす．なぜなら，Max pooling はフィルターの一つでも大きな値があれば全てその値になる処理であるためである．太線化が行われることで，筆跡の線がずれていてもそのずれが結果に影響せず，細かい線の特徴によってうつ病の程度を予測することが難しくなる．この問題に対処するため，Max pooling 層を削除した．

さらに性能を向上させるために，ResNet を用い，スキップ接続を使うことで，筆跡画像特徴を取得する．

3.3 筆跡速度と注目領域の関係

モデルを検証するために，筆跡速度とモデルの関心領域との関係を調べた．先行研究では，うつ病患者は健常者よりもタスクを完了するのに時間がかかる傾向があることが示されている．モデルが手書きスピードの遅い領域に注目していることが示せれば，うつ病患者が認知的に注意を必要とする領域，すなわち書き始めと書き終わりに注意が向けられていることを示すことができる．これを示すために，次のような手法を行った．筆跡速度を時系列データから測定し，正規化して中央値で速いグループと遅いグループに分けた．分母は Grad-CAM++[17]の値が 0.5 以上であり筆跡速度が速いグループの個数を表し，分子は Grad-CAM++の値が 0.5 以上で筆跡速度が遅いグループの個数を表す．そして，これら 2 つの値の比を計算する．比率が 1 を超える場合，このモデルは手書きスピードが遅い領域により焦点を当てていることを示す．上記の方法を検証することで，先行研究で示されたように，モデルが先行研究の特徴を示すことが実証される．

4. 実験

4.1 目的

本実験の目的は 2 つある．1 つ目は筆跡画像に基づくうつ病重症度推定の精度を評価し，提案モデルの有効性を検証することである．2 つ目は Grad-CAM++で強調された領域と筆跡速度の関係を分析し，提案モデルが先行研究で確認された特徴を捉えているかを検証することである．

4.2 データセット

EMOTHAW データセットは 129 人の参加者からなり，全員がうつ病の重症度を評価するために DASS テストを受けた．テストのスコアは 0～25 点で，被験者の内訳は， 95 人が 0～9 点，14 人が 10～13 点，13 人が 14～20 点，7 人が 21 点以上である．参加者はタブレットを使って 7 つのタスク（5 つの描画と 2 つの筆記）に取り組み，筆跡の時系列データが収集された．各タスクについて，x 座標，y 座標，タイムスタンプ，ペンが紙に接地しているかどうか，経度，緯度，筆圧の 8 種類の時系列データが記録された．本研究では，画像としてモデルを学習させるために，x 座標，y 座標，タイムスタンプの 3 つの特徴を使って画像を作成した．

このデータセットは，被験者間のうつ病の重症度分布が重症度の軽い方に偏っている．例えば，最も頻度の高い重症度は 16 人の被験者がいるラベルは 4 であり，16×4×7=448 枚の画像が得られる．このアンバランスに対処するため，各重症度ラベルに 448 枚の画像が含まれるように，データの増強を行った．この増強には，平行移動や縮小などの技法が用いられ，合計 11,200 枚になった（ラベル 11 は存在しない）．

完全なデータセットはその後，3 つの異なる方法で訓練とテストのサブセットに分割した．

1 つ目のデータセットは 129 人の参加者を 9 : 1 の割合で分け，116 人の参加者のデータをトレーニング用に，13 人の参加者のデータをテスト用に使用した．テストに選ぶ 13 人は，偏りを避けるために，集合{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 21, 25}から 13 個のラベルをそれぞれ 1 人ずつ代表するように選ぶ．1 つ目の分割方法を取る理由は，将来の応用のために最も実用的で現実的なアプローチを表しているからである．私たちの目標は，過去のデータでモデルを訓練して，新しい画像を用いて，うつ病の重症度を予測できるようにすることである．したがって，このデータセットは，この現実的な使用事例に最も近い分割方法を反映している．

2 つ目の分割方法は，各参加者が作成した 4 つの手書き単語を 2 つのグループに分け，3 つの単語をトレーニングデータ，1 つの単語をテストデータとして使用した．2 つ目の分割方法をとる理由は，モデルが個人の書き癖を学習できるかどうかを評価することである．同一人物が書いた 4 つの単語をトレーニング用に 3 つ，テスト用に 1 つ分けることで，モデルが 3 つの単語から学習した筆跡パターンを基に，残りの単語からうつ病の重症度を予測できるかどうかを検証する．

3 つ目のデータセットは，完全にランダムに 3:1 で訓練データ，テストデータに分割した．3 つ目の分割方法を取る理由は，オーギュメンテーションした画像も含まれている中，モデルが適切に推測できているかを確認するためである．

4.3 方法

上記の目的を達成するために，2 つ行った．1 つ目はデータを用いて，うつ病重症度推定モデルを学習することである．そして，MSE を用いて精度を評価し，提案モデルを検証する．その時に，実際のうつ病重症度スコアと予測値を箱ひげ図に図示する．散布図では，データ数が多く，特徴を見ることが困難だからである．

2 つ目は特徴の検証である．先行研究で特定された特徴（うつ病患者はタスクを完了するのに時間がかかる）を検証するために，手書き速度と Grad-CAM++によって強調された領域との関係を分析する．

表 1 ピクセル比率を求めるためのパラメータ

G	各画素の Grad-CAM++値．0 から 1 の範囲で，モデルがその画素に払う注目の度合いを表す
v_s	手書きスピードが遅いデータポイントのグループで，スピードの中央値に基づいてデータセットを分割することによって決定される
v_f	手書きスピードが速いデータ点のグループで，スピードの中央値によって決定される
$N_{v_s}(G > 0.5)$	$G > 0.5$であるv_s内のピクセルの総数
$N_{v_f}(G > 0.5)$	$G > 0.5$であるv_f内のピクセルの総数

rをv_s, v_f間の$G > 0.5$であるピクセル比率だとすると，rの式は次のようになる．

$$r = \frac{N_{v_s}(G > 0.5)}{N_{v_f}(G > 0.5)}$$

この式は，遅いグループv_sの$G > 0.5$のピクセルの，速いグループv_fと比較した相対的な注意率を表している．xの値が大きいほど，v_fに比べv_sに$G > 0.5$の画素が集中していることを示す．したがって，rの値が大きいと，学習モデルが遅い領域に焦点を当て，先行研究で示されたうつ病の特徴を反映していることが示唆される．

4.4 結果

データセット 1

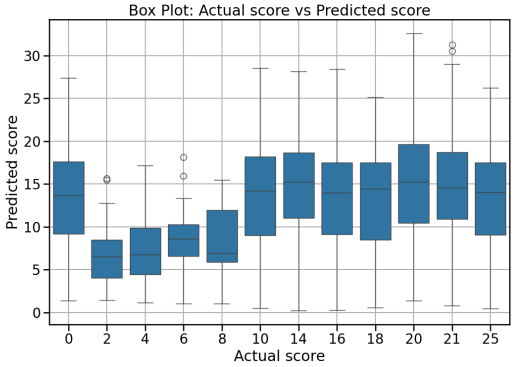

図 4 データセット 1 の結果

トレーニング中のモデルの平均二乗誤差は 1 以下であった．最終的な平均二乗誤差（MSE）の値は約 68 となった．上の箱ヒゲ図に示されているように，モデルは識別可能な傾向を示しており，基本的なパターンをある程度学習したことを示唆している．

データセット 2

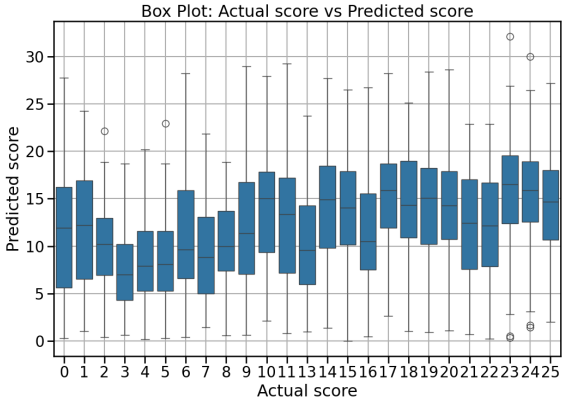

図 5 データセット 2 の結果

最終的に MSE の値は約 68 となった．被験者の書き癖を学習すれば精度が向上すると期待されたが，MSE はデータセット 1 とほとんど同じであった．

データセット 3

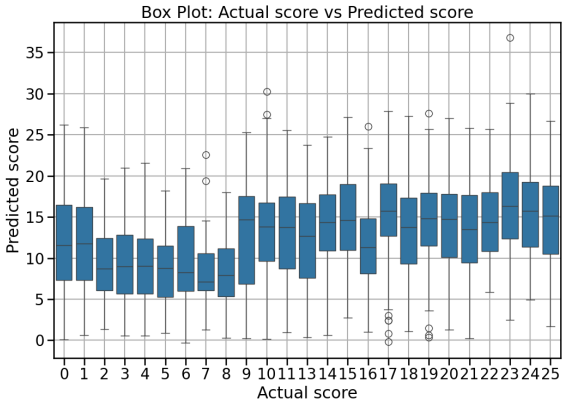

図 6 データセット 3 の結果

最終的な MSE の値は約 65 になった．データオーギュメンテーションを行った画像があるため多少精度が上がったものの，MSE が 0 に近くなるような結果にはならなかった．

筆跡速度と Grad-CAM++の注目領域の関係

図 7 筆跡速度を表した図(上)と Grad-CAM++の画像(下)

上図は，筆跡の速度が速い時を赤色，遅い時を青色で描いた図である．下図は Grad-CAM++の注意領域を白で表した時の画像である．

提案手法で説明した方法に基づいて，先行研究の特徴を学習し，手書き速度が遅い領域についてモデルがより着目しているかどうかを検証する．分母は，Grad-CAM++値が 0.5 以上の速いグループの数を表し，分子は，Grad-CAM++値が 0.5 以上の遅いグループの座標の数を表す．この値をrとすると，$r = 1.21$である．この結果は，Grad-CAM++が手書き速度が遅い領域により焦点を当てていることを示している．

4.5 考察

本研究では，手書きデータを用いてうつ病の重症度をある程度予測できることを示した．また，データセット 2, 3 の結果から，データオーギュメンテーションによって作られた画像はオリジナル画像から離れた異なる特徴を持ち，精度を上げるために用いることができなかった可能性がある．

図 7 に示すように，Grad-CAM++によって示された注意領域は，筆記の開始時と終了時に集中しているようである．定量的実験により，これらの注意領域は，手書き速度が遅い再領域と関連していることが確認された．これまでの研究で，うつ病患者は健常者と比較して，課題を完了するのに時間がかかることが示されている．このことは，筆記中，筆記速度が遅くなりがちな筆記プロセスの開始時と終了時に，認知機能がより大きく関与していることを示唆している．このことから，我々のモデルはうつ病の特徴を捉えていることが示唆される．

しかし，本研究では，筆圧などの他の要因を考慮していない．さらに，データのサンプル数が限られているため，結果の一般化可能性を評価するためには，さらなる検証が必要である．

5. おわりに

手書き文字からうつ病の重症度を推定するモデルを開発し，ある程度予測できることを確認した．ResNet と Grad-CAM++を利用し，注目領域と手書きスピードの関係を調べたところ，モデルは主に遅いスピードに関連する特徴を学習することが確認された．このことは，モデルが先行研究で特定された特徴を捉えていることを示唆している．本研究は，時系列データを画像として保存して推論を行うことの実現可能性を示すものであると考える．

本研究では，画像に焦点を当てたが，今後の研究では，さらなる特徴を探求し，精度の向上を目指す予定である．また，データセットはイタリアで収集されたものであることに留意することも重要である．したがって，他の国の個人のデータを使用して同様の結果が得られるかどうかを調査することが重要である．

参 考 文 献

[1] World Health Organization, “Depression and other common mental disorders: global health estimates,” 2017.

[2] D. A. Regier et al., “DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses,” American Journal of Psychiatry, vol. 170, no. 1, pp. 59–70, 2013.

[3] A. Esposito and A. M. Esposito, “On the recognition of emotional vocal expressions: motivations for a holistic approach,” Cognitive Processing, vol. 13, no. 2, pp. 541–550, 2012.

[4] J. R. Williamson et al., “Vocal biomarkers of depression based on motor incoordination,” Proc. of the 3rd ACM Int. Workshop on Audio/Visual Emotion Challenge, pp. 21–28, 2013.

[5] M. Valstar et al., “AVEC 2013: The continuous audio/visual emotion and depression recognition challenge,” Proc. of ACM Int. Workshop on Audio/Visual Emotion Challenge, pp. 3–10, 2013.

[6] M. Valstar et al., “AVEC 2014: 3D dimensional affect and depression recognition challenge,” Proc. of ACM Int. Workshop on Audio/Visual Emotion Challenge, pp. 3–10, 2014.

[7] X. Zhou, K. Jin, Y. Shang, and G. Guo, “Visually interpretable representation learning for depression recognition from facial images,” IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 542–552, Jul.–Sep. 2020.

[8] X. Ma, D. Huang, Y. Wang, and Y. Wang, “Cost-sensitive two-stage depression prediction using dynamic visual clues,” Proc. of Asian Conference on Computer Vision, pp. 338–351, 2017.

[9] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, “Embodiment of sadness and depression-gait patterns associated with dysphoric mood,” Psychosomatic Medicine, vol. 71, pp. 580–587, 2009.

[10] J. Z. Canales, J. T. Fiquer, R. N. Campos, M. G. Soeiro-de-Souza, and R. A. Moreno, “Investigation of associations between recurrence of major depressive

disorder and spinal posture alignment: A quantitative cross-sectional study," Gait & Posture, vol. 52, pp. 258–264, 2017.

[11] A. Seal et al., "DeprNet: A deep convolution neural network framework for detecting depression using EEG," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[12] I.-M. Spyrou et al., "Geriatric depression symptoms coexisting with cognitive decline: A comparison of classification methodologies," Biomedical Signal Processing and Control, vol. 25, pp. 118–129, Mar. 2016.

[13] A. Seal et al., "DeprNet: A deep convolution neural network framework for detecting depression using EEG," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[14] L. Likforman-Sulem et al., "EMOTHAW: A novel database for emotional state recognition from handwriting and drawing," IEEE Transactions on Human-Machine Systems, vol. 47, no. 2, pp. 273–284, 2017.

[15] J. A. Nolazco-Flores et al., "Emotional state recognition performance improvement on a handwriting and drawing task," IEEE Access, vol. 9, pp. 1–10, 2021.

[16] K. He et al., "Deep residual learning for image recognition," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[17] A. Chattopadhay et al., "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 839–847, 2018.