

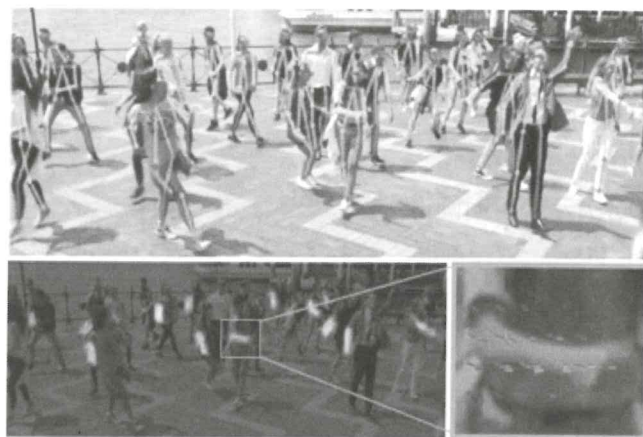
ーラルネットワーク (RNN^{※9}) や時系列向けのニューラルネットワークであるLSTMなどの手法も使われる。どちらもネットワーク内にフィードバック結合を持ち、時系列データの処理中に状態を記憶し続けることができる。LSTMは、スマートフォンに搭載された音声認識機能で使われ^{※10}、一般向けのシステムとしては高い精度を持ち、現在では、人間以上の認識率を持つシステムも登場している。

その他の認識分野

画像や動画に対する認識処理が可能になってくると、人体の姿勢推定、人物の識別 (顔だけでなく体格とその動きの特長など)、人の感情の推定も可能になってきた。

人物の姿勢推定は、人の頭部、肩、肘、手、腰、膝、足を検出し、関節のつながりなどから、人がどのような姿勢を取っているかを推定する技術である。OpenPose (図2-2-5) は、カーネギーメロン大学 (Carnegie Mellon University、米国) のZhe CaoらによってCVPR2017 (Conference on Computer Vision and Pattern Recognition) で発表された、ディープラーニングの適用で複数人の姿勢を同時に推定するアルゴリズムが実装されたライブラリである。画像から各関節位置を、ディープラーニングを用いて推定し、推定された関節をつなぎ合わせることで姿勢を推定する。

■図2-2-5 OpenPoseの例



出典: Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields^{※11}

人の感情については、顔の動きをコード化することが可能 (例「FACS ; Facial Action Coding System」など) で、そのコードと感情の関係などが研究されてきた。映像からコードを認識することで、感情を推定することも可能になってきた。活動時の人間の感情を認識することで、安全などのために適切な動作を行うこと (例えば自動車運転者の状態監視) や、質問回答時の言葉に表れない感情の動きなどを推定することで、消費者の反応を得るといった応用分野が考えられる。また、行動認識とスポーツの組み合わせでは、選手の動作の検証や判定、あるいは採点といった応用分野が考えられる。

※9 RNN (Recurrent Neural Network) はディープラーニングのネットワーク構造の一つ、時系列データの予測に優れているとされる (詳細は2.3参照)。

※10 Google voice search: faster and more accurate, Google Blog, 2015 <<https://ai.googleblog.com/2015/09/google-voice-search-faster-and-more.html>>

※11 <https://arxiv.org/pdf/1611.08050v2.pdf>

(3) 最新技術動向

隠れた物体認識 (NOLS、GQN)

ディープラーニングでは、データの中にあっても人が気づいていない、あるいは直接見えないパターンを特徴量として利用することができる。例えば、画像認識では、何かに遮られて直接見ることができない位置にある物体を、可視範囲の別の物体からの反射光などを利用して認識するNOLS (Non Line of sight) も登場している。これらは自動運転や遠隔センシング、医療、監視といった様々な分野での応用が期待され、CVPR2019でも6本の論文 (うち1本がbest Paper) が採録された。

また、写真などの2次元画像から、被写体となった3次元物体を推定するGQN (Generative Query Network) も登場している。GQNでは、カメラを移動させ、その画像から、3次元モデルを生成していく。これには、スマートフォンのような機器で、その位置や動きを正確に測定しつつ、画像を撮影できるようになった点が大きい。

顔認識 (DeepFaceなど)

人間の顔の認証もかつては、特徴量の抽出と判定を組み合わせていたが、現在では、ディープラーニングを利用するものが主流である。2015年に公開されたFacebookのDeepFaceは、Facebookにアップロードされていた400万人のイメージを使い、9層のネットワーク、1億2,000万の重み係数を持つもので、97.35%という高い認識率を示した。物体認識同様、大量の学習、テスト画像が利用できるようになった点が大きな理由だ。

顔認識では、基本的なデータセット、例えば既存の物体認識の学習済みモデル (AlexNetなど) を活用し、これに活性化関数、損失関数を組み合わせて、全体として再学習を行うことで特定の人物かどうかの判定ができる。また、最初から顔認識に特化したディープニューラルネットワークとしてFace++、MageFace、FaceNetなどがある。

文字認識 (AI-OCR)

手書きや印刷物の文字認識技術であるOCR (Optical Character Recognition) にもディープラーニングの波が押し寄せている。ディープラーニングなどのAI技術を利用したOCRは、AI-OCRなどと呼ばれる。AI-OCRでは、認識精度が上がっただけでなく、従来、認識時に設定する必要があったレイアウト情報 (どの欄に住所や氏名が書かれているかといった情報) がなくても自動的に判別し、異なる書式の伝票から住所や氏名などの項目を読み出して整理することができるようになった点が大きい。