

携帯型GPUスパコン フル回転! AI ツアー・ガイドを作る

村上 真奈, 矢戸 知得



説明文が表示され、音声でも伝えてくれる



説明文：a black and white cat sitting
in a bathroom sink

(洗面所のシンクに白と黒の猫が座っている)



説明文：a couple of people
that are playing a video game

(テレビ・ゲームを遊んでいる2人の人)

写真1 作るもの…カメラで外の世界を見て人工知能が理解したことを音声で伝えてくれる携帯型音声ガイドAIデバイス

囲碁対局における「AlphaGo」の快挙などから分かるように、目覚ましい速度で人工知能(AI: Artificial Intelligence)の研究の成果が上がっています。ディープ・ラーニング(Deep Learning; 深層学習)用のフレームワークも公開され、その活用にも注目が集まっています。

ここでは、これらの技術的な進歩を利用して、手で動く面白い人工知能デバイスを作ってみます。

作るもの

● 携帯型音声ガイドAIデバイス

ここでは、カメラで外の世界を見て、人工知能が理解したことを音声で伝えてくれる音声ガイド・デバイスを作ります(写真1)。屋外で利用できるように、電池で動作し、持ち運べるものを目指します。

このデバイスからの音声ガイドを聞いて街を歩き回ることができるかもしれません。視覚に障害を持った人に対して補助をするデバイスや、観光名所・地点のデータを学習させておくことでツアー・ガイド装置として応用することも考えられます。定期的に学習済みのモデルだけを更新することで、新しい人気スポットや注意すべき工事地点などへの対応も効果的に行うことができるでしょう。

ディープ・ラーニング・アプリケーションと学習済み

のモデルを持って、認識処理の全てをデバイス上で行うことも特徴です。ネットワーク越しにクラウド・サービスに投げて、その判定をサーバ上で行う訳ではありません。そのため遅延が少なく、ネットワークに接続できない場所での利用も考えられます。

● 未知の画像から説明文を生成する

製作につなげる話題として、一つの研究成果を紹介しましょう。

ディープ・ラーニングを利用して、コンピュータは人間以上の精度で物体を認識する能力を示すようになってきました。この分野の研究はさらに進んで、より複雑なニューラル・ネットワーク(Neural Network)を組み合わせることで、写真の中の物体の関係まで理解できるようになってきています。

その一つが、画像データと言語の関係を学んだニューラル・ネットワークが、未知の画像に対して、人が理解する英語の説明文を生成する研究です⁽¹⁾(図1)。人工知能の進化について語った動画⁽²⁾でも紹介され、「機械がここまで人と同じように物事を理解できるようになっているのか」と非常に大きなインパクトを与えました。

● コードもモデルも公開されている

まさに最先端の研究成果ですが、このモデルを