

## VaDE-GAN と Grad-CAM を用いた

## HMD 作業支援システムに対する支援挙動生成手法

## Automatic Generation Method of Support Behavior for HMD Work Support System based on VaDE-GAN and Grad-CAM

橋本 幸二郎<sup>†</sup>      三代沢 正<sup>‡</sup>      山田 哲靖<sup>‡</sup>  
 Kohjiro Hashimoto    Tadashi Miyosawa    Tetsuyasu Yamada

## 1. 背景と目的

人中心の作業を担う中小企業において、少子高齢化に伴う人材不足が深刻な問題となっている。対して、Head Mounted Display(HMD)が作業を支援するツールとして注目を集めている。HMD は作業者の視界上にデジタル情報を重畳することができ、リアルタイムな情報提示が可能となる。それ故、適切なタイミングで適切な支援情報が提示されるよう設計することにより、初心者でも複雑な作業をこなすことが期待できる。

しかし、支援すべき作業の種類が多い場合、その一つ一つの作業に対して支援挙動を設計する必要がある。この場合における設計者の負担は計り知れず、システムの支援挙動を自動で設計する技術の開発が必要と考える。

本研究では、ノート PC の修理作業を例に挙げる。計算機の修理を担う工場では、数十種類以上の計算機を扱う。さらに、人材不足の影響により、その修理作業を一人で担う企業も存在する。この場合、作業には熟練が必要となり、初心者では担うことは難しい。そこで、本研究では、ノート PC の修理作業を対象とした HMD による作業支援システムを想定し、その作業支援挙動を自動で設計する手法を提案する。

## 2. HMD 作業支援システムの概要

図 1 に本研究で想定する HMD 作業支援システムの概要図を示す。作業者は HMD を装着しながら作業を行う。作業中の HMD 搭載カメラ映像と別途設置された固定カメラ映像はサーバへ送られ、サーバでは作業の種類、工程等の認識処理が行われ、最終的に支援情報が決定する。その支援情報は HMD へ送られ、作業者の視界上に提示される。

本研究の課題は、サーバで行われる、支援情報を決定するまでの一連の処理を自動で設計する事である。ここでは、支援情報の決定処理を行う数理モデルを支援挙動モデルと呼ぶ。そして、作業中の映像から支援挙動モデルを自動生成する手法を検討する。本稿では問題を単純化するため、以下の条件における手法の検討を行った。

一つは、作業内容はノート PC の部品取り外し作業とする。こうすることで、作業者が取るべき行動を”取り外す”に絞る。二つ目は、HMD より提示する支援情報は作業箇所とする。すなわち、作業者には取り外す箇所が提示される。最後に、本稿では固定カメラの映像のみを用いて手法を検討する。すなわち、HMD 搭載カメラ映像は本稿では

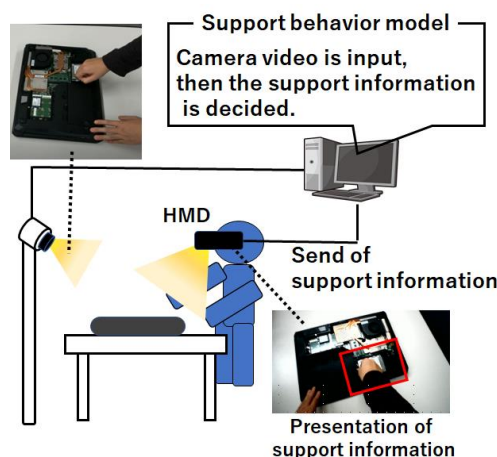


図 1：想定する HMD 作業支援システムの概要

用いないことを意味する。これは、作業者の移動や姿勢の変化に伴い、カメラ外部パラメータが大きく変動するため、3次元物体認識等の高度な技術が要求されることが想定されるためである。

提案手法では、深層学習技術である VaDE-GAN(Variational Deep Embedding-Generative Adversarial Networks)と Grad-CAM を組み合わせたモデルアーキテクチャを提案する。この提案モデルでは、作業工程モデル及び作業箇所の検出モデルを作業映像のみから自動生成できることが期待できる。そこで本稿では、実験を通じて、提案モデルの有効性を検証した。

## 3. 提案手法

図 2 に支援挙動モデルを表現する深層学習モデルアーキテクチャを示す。提案モデルは、物体検出部、作業工程モデリング部、作業箇所検出部で構成されている。次節に詳細を述べる。

## 3.1 物体検出部

物体検出部は Single Shot MultiBox Detector(SSD)[1]で表現され、作業映像からノート PC を検出する処理を行う。SSD は、Microsoft が提供する MSCOCO データの学習済みモデルが公開されている。そのモデルにはノート PC 画像が学習済みであり、本稿では、この学習済みモデルを用いることによりノート PC を検出する。SSD により検出されたノート PC の領域画像は、作業工程モデリング部である VaDE-GAN へ入力される。

<sup>†</sup> 公立諏訪東京理科大学 工学部情報応用工学科  
 Suwa University of Science, Department of Applied  
 Information Engineering

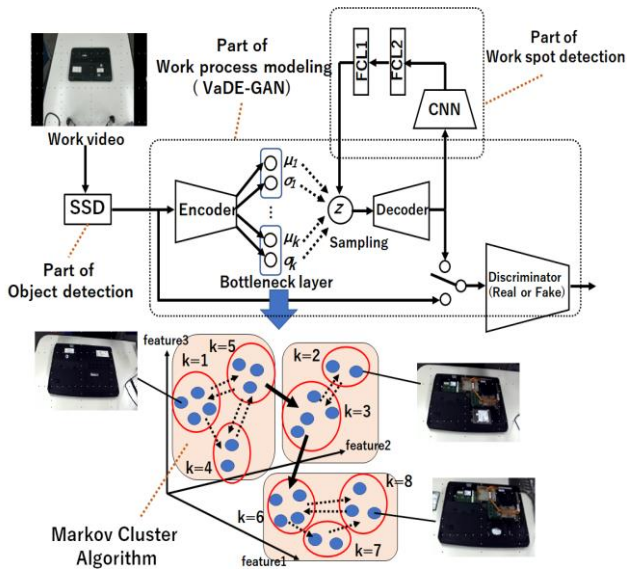


図2：提案モデルのアーキテクチャ

### 3.2 作業工程モデリング部

作業工程モデリング部は、VaDE[2]とGAN[3]を組み合わせたモデル(VaDE-GAN)で表現される。VaDEは、画像生成モデルの一つであり、ボトルネック層にて混合ガウスモデルが学習されることから、入力画像に対する低次元特徴量化及びクラスタリングを同時に行うことができる。さらに、このモデルにGANを組み込み、VaDEより生成される画像をより鮮明にする。さらに、提案モデルでは、図2下図に示すように、クラスタ間の遷移に対して、Markov Cluster Algorithmを適用することにより、作業工程を表す状態遷移ネットワークを生成する。

### 3.3 作業箇所検出部

作業箇所検出部は、作業工程の各状態を識別するCNNとCNNの出力をVaDEの潜在変数に変換する全結合層で表現される。CNNに対してGrad-CAM[4]を適用することにより、作業工程の各状態を識別した際の要因領域を抽出する。ここでは、Grad-CAMにより抽出される領域を作業箇所とみなす。

### 3.4 作業箇所検出の流れ

図2のアーキテクチャでは、作業映像を学習データとして、End-to-Endに作業工程モデリング部及び作業箇所検出部が学習される。そして、生成されたモデルに対して、次の流れで作業箇所の検出及び提示を行う。

まず、作業映像はSSDに入力され、ノートPC画像領域が検出される。検出された領域画像はVaDE-GANへ入力され、その生成画像がCNNへ入力される。CNNでは現在の作業工程を認識する。さらに作業工程を表す状態遷移ネットワークに基づき、次の工程を表すOne-hotベクトルを生成する。生成したOne-hotベクトルはVaDE-GANの潜在変数へ変換され、VaDE-GANによる画像生成を経て、次の作業工程を表す画像を生成する。最終的に、次の作業工程を表す画像がCNNへ入力され、Grad-CAMに基づき識別要因領域を作業箇所領域とみなして検出される。

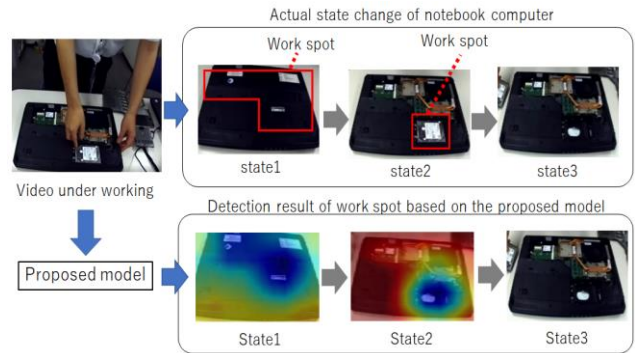


図3：提案モデルに基づく作業箇所検出結果例

## 4. 評価実験

提案手法の有効性を確認する実験を行った。修理用ノートPCを3台用意し、それぞれに対して、Battery取り外し、HDD取り外し、FAN取り外しの3種類の作業を支援対象とした。提案手法に基づきそれぞれの支援挙動モデルを生成し、状態の認識精度及び作業箇所検出精度を評価した。

図3にある修理用ノートPCのHDD取り外し作業における作業箇所の検出結果例を示す。HDD取り外し作業においてノートPCは、作業前の状態(State1)、カバー取り外し状態(State2)、HDD取り外し状態(State3)と状態を遷移する。図3(上図)が各状態に対する実際の作業箇所を示し、図3(下図)が提案モデルに基づき検出された作業箇所をヒートマップで示している。両者を比べると提案モデルに基づき作業箇所とみなされる領域が検出されていることが確認できる。この時の状態認識精度及び作業箇所の検出精度を定量的に評価しており、その詳細については当日発表する。

## 5. まとめ

本稿では、HMD作業支援システムを設計する上での、支援挙動を自動生成する手法を検討した。深層学習技術であるVaDE-GANとGrad-CAMを利用し、作業工程のモデリング及び作業箇所検出モデルをEnd-to-Endで学習可能なモデルアーキテクチャを提案した。提案モデルに対する定量的評価結果の詳細については当日に発表する。

### 謝辞

本研究は、鷹野学術振興財団及び高橋産業経済研究財団の助成を受けたものである。ここに謝意を示す。

### 参考文献

- [1] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, SSD: Single Shot MultiBox Detector, In European Conference on Computer Vision, arXiv:1512.02325v5, pp.21-37 (2016).
- [2] Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou, Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering, proceedings of the 26<sup>th</sup> International Joint Conference on Artificial Intelligence, pp.1965-1972 (2017).
- [3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil, Aaron Courville, Yoshua Bengio, Generative Adversarial Nets, proceedings of Neural Information Processing Systems, pp.2672-2680 (2014).
- [4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, proceedings of International Conference on Computer Vision, pp.618-626 (2016).