

# 半教師あり学習によるアイコンタクト検出

三鼓 悠<sup>1,a)</sup> 中澤 篤志<sup>1</sup> 西田 豊明<sup>1</sup>

**概要：**本論文では半教師あり学習によりアイコンタクトを検出する手法を提案する。我々は、人同士のコミュニケーション解析のためのアイコンタクト識別器を深層学習技術を用いて実現してきたが、それには必要となる訓練データを作成するコストが高いという問題がある。本研究では Generative Adversarial Network (GAN) を用いた半教師あり学習をアイコンタクト検出器の学習に用いることにより、少数の教師あり訓練データにより検出器を作成する。実験では学習に用いる教師あり訓練データ数を変え比較した。その結果、半教師あり学習がアイコンタクト検出器の作成に有効で従来よりも少ないデータ数で識別精度 80% を達成することができた。一方で目領域画像のみからアイコンタクトを行う性能的な限界についても議論する。

**キーワード：**アイコンタクト, 一人称視点映像, 半教師あり学習

## Eye contact detection with Semi-supervised learning

YU MITSUZUMI<sup>1,a)</sup> ATSUSHI NAKAZAWA<sup>1</sup> TOYOAKI NISHIDA<sup>1</sup>

**Abstract:** In this paper, we show a method to detect eye contact by using semi supervised learning. In our previous study, we developed eye contact detection algorithms for human-to-human communication analysis by using deep learning technology; however it required a lot of efforts for making annotated learning data. To solve them, We introduce semi supervised learning using GAN for training eye contact detector, created it with a small number of supervised training data. In the experiment, We compare several patterns of the number of supervised training data. As a result, we find that semi supervised learning is effective for creating eye contact detector and it achieves accuracy of 80% with a smaller number of data than before. On the other hand, we also discuss the performance limits of eye contact from eye area image alone.

**Keywords:** Eye contact, First person view video, Semi-supervised learning

### 1. はじめに

アイコンタクトとは視線を合わせることであり、非言語的なコミュニケーションの1つである。アイコンタクトを検出することは健常者のコミュニケーションの解析に有益である。

自閉症スペクトラム (ASD) の人はアイコンタクトが困難な割合が多い。また乳幼児の視線情報には ASD の早期発見に有用であるという報告がある [1] ことから、ASD の早期発見のためにアイコンタクトの成否を利用することも

ある。

また、認知症ケアスキルの1つとしてもアイコンタクトは重要視されており [2], 介護者の一人称視点映像からアイコンタクトを評価する研究 [3] が行なわれている。

これらの観点から我々はアイコンタクトを機械学習により識別するシステムについて研究を行ってきたが、学習に必要なデータセットを作成することはコストが高いという問題がある。またデータセットが用意できたとしても例えば介護現場で利用する場合、データセット中の人物の年齢と、対象者の年齢との差異などからうまく適用できないという問題も存在する。

本研究ではアイコンタクト識別器の生成に近年研究が盛

<sup>1</sup> 京都大学, 〒 606-8501 京都府京都市左京区吉田本町

<sup>a)</sup> mitsuzumi@ii.ist.i.kyoto-u.ac.jp

んに行われている Generative Adversarial Network (GAN) [4] を用いた半教師あり学習手法を取り入れることで、データセット作成コストの削減とそれによって多様な場面での識別器の運用を可能とすることを目的とする。

## 2. 関連研究

機械学習を用いアイコンタクト検出器を生成する試みはこれまで様々行われてきた。Smith ら [5] は顔全体から両目の縁を検出することで目画像を得、特徴抽出し、主成分分析と多変量解析により次元圧縮したのちに SVM で識別を行う。Ye ら [6] は顔画像を顔方向で 3 つにクラスタリングし、それぞれのクラスタごとの RandomForest 分類器に目画像を入力することで識別を行う。我々は両目画像を入力として CNN を学習することで識別器を生成する研究を行ってきた [7]。これらの手法は全て、教師ありデータを用いて識別器を学習する手法である。

また Zhang らの手法 [8] によるアイコンタクト検出では、全顔画像と顔ランドマーク特徴から CNN 特徴量を抽出し SVM を学習することで生成する。ここではデータセットに対してアノテーションをつける際に彼らの視線推定の手法 [9] を用いて 2 次元平面状に視線情報をプロットし、クラスタリングをクラスタ毎にアイコンタクトのアノテーションをつける方法を採用している。

これらの手法は Zhang らの手法以外は全てデータセットを作成する必要がある、また Zhang らの手法も視線推定を行うための学習用データセットが必要になり、いずれの手法においてもデータセットを作成する際のコストが課題になる。本研究の目的であるデータセット作成にかかるコストの削減はこれらの課題解決に有効であると考ええる。

## 3. 手法

図 1 に半教師あり学習によるアイコンタクト検出器の学習方法、図 2 にアイコンタクト識別の概要を示す。図に示すとおり本手法は顔画像から目画像を抽出するモジュールと、目画像を入力としてアイコンタクトを識別するモジュールの二つからなる。

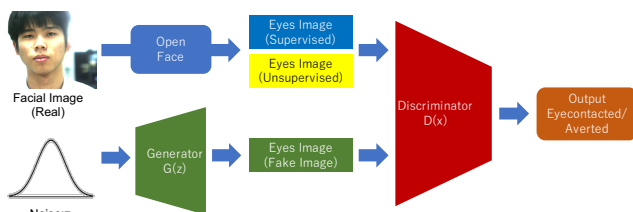


図 1 学習フェイズ

### 3.1 目画像抽出

本手法ではまず入力として顔全体を含んだ画像を受け取

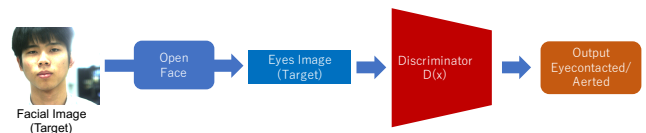


図 2 識別フェイズ

る。入力画像に OpenFace[10] を適用し、入力画像に含まれる顔のランドマーク点を 68 点を取得する。取得したランドマーク点のうち目領域抽出に関係するランドマーク点は片目につき 6 点得られる (図 3 左、黄色点)。片目周辺のランドマーク点をそれぞれ  $l_1, l_2, \dots, l_6$  とすると抽出される領域は

$$\begin{aligned} x_1 &= \min_i l_{ix}, y_1 = \min_i l_{iy} \\ x_2 &= \max_i l_{ix}, y_2 = \max_i l_{iy} \\ w &= x_2 - x_1, h = y_2 - y_1 \end{aligned}$$

としたとき以下で得られる矩形  $(x'_1, y'_1, x'_2, y'_2)$  である (図 3 左、青枠)。

$$\begin{aligned} x'_1 &= x_1 - 0.1w, y'_1 = y_1 - 0.1h \\ x'_2 &= x_2 - 0.1w, y'_2 = y_2 + 0.1h \end{aligned}$$

すなわち、目の周辺のランドマーク点を全て含む最小矩形について 1.2 倍の大きさの領域を抽出対象とする。

得られた目領域画像をグレースケール化し左右の目画像を結合したもの (図 3 右) が次節の識別器の入力となる。

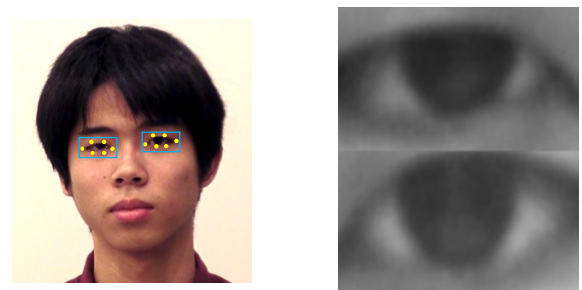


図 3 左: 目周辺のランドマーク点 (黄色) 及び抽出対象領域 (青色)、右: 識別器入力画像

### 3.2 識別器

本研究ではアイコンタクトの識別器を Improved-GAN[11] の手法を用いて生成する。

#### 3.2.1 ネットワーク構造

ネットワークは Generator と Discriminator の二つからなる。それぞれのネットワーク構造を図 4, 5 に示す。これらのネットワークは DCGAN[12] をベースとした。

Discriminator は 4 つの畳み込み層と 1 つの出力層からなる。畳み込み層のフィルタサイズは  $5 \times 5$ 、フィルタの移動幅は 2 である。各畳み込み層では Batch Normalization を適用した。また各畳み込み層には活性化関数とし

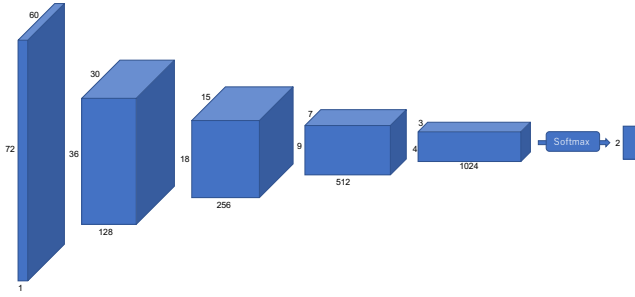


図 4 Discriminator のネットワーク構造

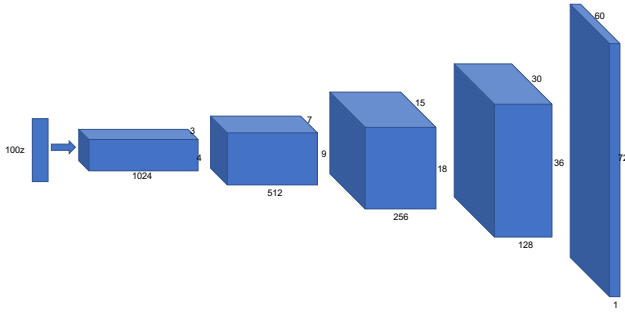


図 5 Generator のネットワーク構造

て Leaky ReLU を適用した。出力層は通常の GAN における Discriminator とは異なり、入力された目画像がアイコンタクトしているか否かの識別を行うため、出力サイズが 2 となっており、活性化関数は Sigmoid 関数の代わりに Softmax 関数を適用する。

Generator はサイズが異なる点を除けば DCGAN における Generator と同様のものを用いており、逆畳み込み層 4 つからなる。各層の活性化には出力層以外では ReLU 関数を、出力層では Tanh 関数を用いる。

### 3.2.2 学習フェイズ

GAN の学習と同様に Discriminator は実在画像と生成画像を識別するように、Generator は Discriminator が実在画像と識別するような画像を生成するように学習させる。目的関数は以下ようになる。

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

ここでクラス数を  $K = 2$ 、すなわちアイコンタクトの有無の 2 クラスとすると、Discriminator の損失関数は

$$\begin{aligned} L_D &= -\mathbb{E}_{x, y \sim p_{data}(x, y)} [\log p_{model}(y|x)] \\ &\quad -\mathbb{E}_{x \sim G} [\log p_{model}(y = K + 1|x)] \\ &= L_{supervised} + L_{unsupervised}. \end{aligned}$$

$$L_{supervised} = -\mathbb{E}_{x, y \sim p_{data}(x, y)} \log p_{model}(y|x, y < K + 1).$$

$$\begin{aligned} L_{unsupervised} &= -\mathbb{E}_{x \sim p_{data}(x)} \log [1 - p_{model}(y = K + 1|x)] \\ &\quad -\mathbb{E}_{x \sim G} \log [p_{model}(y = K + 1|x)] \\ &= -\mathbb{E}_{x \sim p_{data}(x)} \log D(x) \\ &\quad -\mathbb{E}_{z \sim noise} \log(1 - D(G(z))). \end{aligned}$$

であり、教師ありデータからなる  $L_{supervised}$  と教師なしデータ及び生成データからなる  $L_{unsupervised}$  に分けることができる。このうち  $L_{supervised}$  は Softmax クロスエントロピーである。

続いて  $L_{unsupervised}$  については Softmax 関数の性質より

$$D(x) = \frac{Z(x)}{Z(x) + 1}.$$

$$Z(x) = \sum_{k=1}^K \exp[l_k(x)].$$

と表されることから、以下のようにして求められる。

$$\begin{aligned} L_{unsupervised} &= -\mathbb{E}_{x \sim p_{data}(x)} \log D(x) \\ &\quad -\mathbb{E}_{z \sim noise} \log(1 - D(G(z))) \\ &= -\mathbb{E}_{x \sim p_{data}(x)} \{\log Z(x) - \text{softplus}(\log Z(x))\} \\ &\quad -\mathbb{E}_{z \sim noise} \text{softplus}(\log Z(G(z))). \end{aligned}$$

$$\text{softplus}(x) = \log(1 + e^x).$$

Generator の損失関数は以下のようにして求められる。

$$\begin{aligned} L_G &= -\mathbb{E}_{z \sim noise} \log(D(G(z))) \\ &= -\mathbb{E}_{z \sim noise} \{\log Z(G(z)) - \text{softplus}(\log Z(G(z)))\}. \end{aligned}$$

また Feature matching[11] 項を Generator の損失関数に加える。Feature matching 項は実在画像と生成画像の Discriminator の中間層での出力についての平均二乗誤差である。これを最小化することはすなわち実在画像と生成画像の Discriminator における中間特徴を近づけることになり、Generator がより実在画像を生成するように学習を進めることになる。中間特徴として、Discriminator 第 3 層の出力を用いた。

$$R_{feature \text{ matching}} = \|\mathbb{E}_{x \sim p_{data}} f(x) - \mathbb{E}_{z \sim noise} f(G(z))\|.$$

得られた損失関数  $L_D, L_G$  を最小化するように敵対的に学習させることでネットワークを学習させる。ここで最適化アルゴリズムには Adam[13] を用いた。

### 3.2.3 識別フェイズ

アイコンタクトの識別には学習フェイズで生成した Discriminator を用いる。

3.1 節により抽出された目画像を  $x_t$  とすると Discriminator により識別尤度  $P(y|x_t)$  が得られる。ここで閾値  $\tau$  とクラス 1 の尤度について

$$P(y = 1|x_t) > \tau$$

が成立している場合はアイコンタクトしていると識別し、成立していない場合はアイコンタクトをしていないと識別する。

## 4. 評価実験

本手法の識別精度及び半教師あり学習に用いる訓練データ数の検討のための実験を行なった。

### 4.1 データセット

#### 4.1.1 教師ありデータ

教師ありデータとして視線方向がラベルづけ済みの Columbia Gaze Dataset[5]を用いた(図6)。

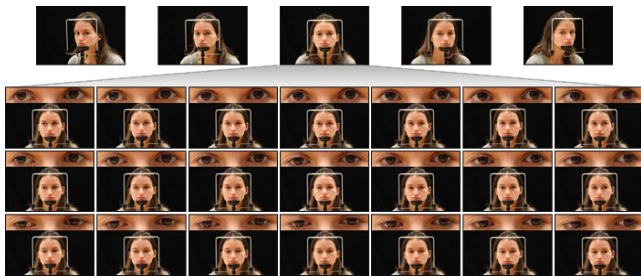


図6 Columbia Gaze Dataset[5]

データセットは撮影方向、視線方向を変化させた56人の顔画像データが5880枚含まれている。撮影方向は水平方向に $0^\circ, \pm 15^\circ, \pm 30^\circ$ の5種類、視線方向は一つの撮影方向につき水平方向に $0, \pm 5^\circ, \pm 10^\circ, \pm 15^\circ$ 、垂直方向に $0^\circ, \pm 10^\circ$ の計21種類含まれる。このうちアイコンタクトしているもの、アイコンタクトしていないものと新たにラベルづけを行い4200枚を実験に用いた。

#### 4.1.2 教師なしデータ

教師なしデータセットとして Facescrub[14]を用いた(図7)。



図7 教師なしデータ: Facescrub[14]

データセットに含まれる人物顔画像から3.1節の方法により目画像を抽出を行い、このうち不鮮明な画像や抽出に失敗した画像など学習に不適切なものを除外し、結果15273枚を実験に用いた。

#### 4.1.3 評価データ

我々は視線方向、撮影方向を変化させて評価用データを撮影した(図8)。

1人につき、視線方向を水平方向 $0^\circ, \pm 30^\circ, \pm 60^\circ$ 、撮影方向を水平方向 $0^\circ, \pm 30^\circ$ に変化させ、合計15種類のデータを撮影した。このような画像を5人の被験者から合計

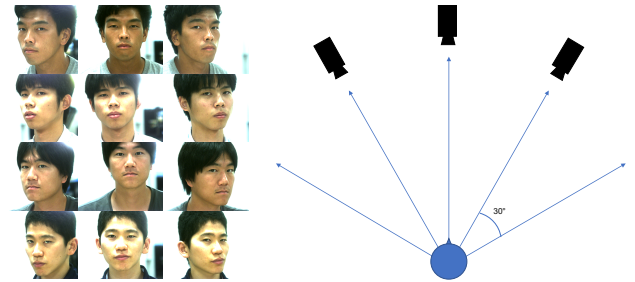


図8 左画像のような評価用データを作成した。右図のように、視線方向を水平方向に $0^\circ, \pm 30^\circ, \pm 60^\circ$ 、撮影方向を水平方向に $0^\circ, \pm 30^\circ$ に変化させて評価データを作成した。

19179枚撮影した。

### 4.2 実験設定

#### 4.2.1 比較対象

実験では以下の手法を比較した。手法の違いについてまとめたものが表である。

**DEEPEC(従来法)** DEEPEC[7]では左右の目を検出、目領域画像を抽出、正規化したものを入力として、2つの並列な畳み込み層を持つニューラルネットワークを教師あり学習により学習することで識別器を作成する。本実験では教師あり学習データには4.1.1節を、評価データには4.1.3節を用いた。

**SSLEC(提案手法)** 提案手法では3.1節により目画像を抽出し、3.2により識別器を作成する。本実験では教師ありデータに4.1.1を教師なしデータには4.1.2を評価データには4.1.3を用いた。また、学習に用いる教師ありデータと精度との関係の評価のために教師ありデータとして用いるデータ数を3360,2100,840に変化させて実験を行なった。

| Methods     | supervised    | unsupervised    | testing    |
|-------------|---------------|-----------------|------------|
| DEEPEC      | 4200 data [5] | -               | 19179 data |
| SSLEC(3360) | 3360 data [5] | 15273 data [14] | 19179 data |
| SSLEC(2100) | 2100 data [5] | 15273 data [14] | 19179 data |
| SSLEC(840)  | 840 data [5]  | 15273 data [14] | 19179 data |

表1 手法比較表

#### 4.2.2 評価方法

評価は各手法によって得られた評価データに対するのアイコンタクトの尤度 $P(y|x_t)$ と閾値 $\tau$ の間に

$$P(y = 1|x_t) > \tau$$

が成立している場合にそのフレームでアイコンタクトが成立している、成立していないとした場合にそのフレームではアイコンタクトをしていないとし、閾値 $\tau$ を変化させることでROC曲線を求め評価する。ROC曲線の評価については曲線下の面積AUC (Area Under the Curve)を算出することによって行う。得られたROC曲線より識別器の



最適閾値  $\tau$  を求め、その閾値による正答率、適合率、再現率、F 値を算出し評価方法とする。

### 4.3 実験結果

各手法の ROC 曲線、正答率、適合率、再現率、F 値を図 4.3 と表 2 にまとめた。

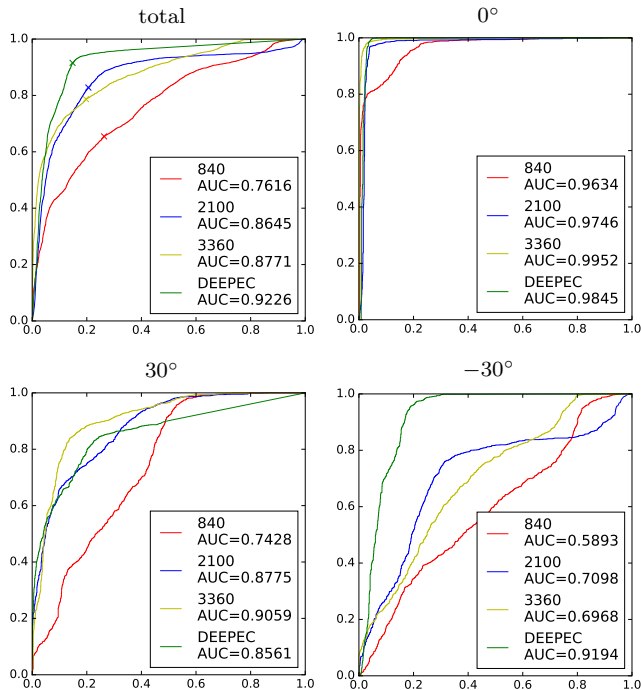


図 9 各手法の識別結果を ROC 曲線で評価した。左上：全撮影方向データに対する識別性能。右上：撮影方向 0° のデータに対する識別結果。左下：撮影方向 30° のデータに対する識別結果。右下：撮影方向 -30° のデータに対する識別結果。

|          | Accuracy | Precision | Recall | F-measure |
|----------|----------|-----------|--------|-----------|
| 3360data | 0.7948   | 0.8006    | 0.7866 | 0.7935    |
| 2100data | 0.8111   | 0.8019    | 0.8277 | 0.8145    |
| 840data  | 0.6954   | 0.7142    | 0.6544 | 0.6830    |
| DEEPEC   | 0.8839   | 0.8617    | 0.9153 | 0.8878    |

表 2 識別結果

全撮影方向データについての ROC-AUC の値、及び F 値について比較を行うと、従来手法が最も良い精度であるという結果が得られた。提案手法における教師ありデータ数と精度の関係について評価した場合、2100 データ用いる場合と 3360 データ用いる場合では精度に差はほぼ生じなかったが、840 データを用いる場合では他 2 つに比べて精度が低いという結果となった。

### 4.4 考察

実験で半教師あり学習により作成したアイコンタクト識別器は、従来の教師あり学習による手法ほどの精度を実現することができなかった。原因として半教師あり学習に



図 10 学習中に Generator により生成された画像。これは教師ありデータを 2100 データ用いて学習した場合の 300 エポック目の生成画像である。

GAN の構造を用いたことがあるのではないかと考える。GAN の学習方法では Discriminator はデータセットの实在画像と Generator が生成した画像をより識別するように学習し、Generator は实在画像により近い画像を生成するように学習を進める。提案手法の Discriminator がアイコンタクトのクラス識別器としての役割と GAN における Discriminator の役割を兼ねており、損失を共有している。したがって Generator で生成された画像を Discriminator が誤識別した場合、それはクラス識別器としての損失にもなる。ここで学習が進んだ結果、図 10 のような Generator が实在画像の分布域のデータを生成するようになった場合について考える。このようなデータを Discriminator が实在画像として識別した場合、GAN の枠組みでは誤識別したとして誤差を逆伝播することになるが、Discriminator のクラス識別器としての役割を考えた場合ではこのようにして学習を進めることは望ましくない。これについては Dai ら [15] が指摘しているが、Generator ができるべく实在画像の分布に近いが分布にはのらないデータを生成するように学習させることで半教師あり学習の精度を上げることができる。この点について、Triple-GAN[16] のような Discriminator とクラス識別器を独立した枠組みを用いる、あるいは Dai らの手法のように Generator が实在画像を生成しないような学習を用いることが必要になると考える。

続いて学習に用いる教師ありデータ数と半教師あり学習の効果について検討を行うと、3360 データ用いる場合と 2100 データを用いる場合では F 値及び ROC-AUC の値についてほぼ同精度であったことから、本手法を用いてデータ数を削減することが有効であるということがわかる。しかしこれら二つの実験結果と 840 データを使用した場合の実験結果では精度に大きく差が生じた。教師ありデータ数

を減らすことにより生じる差異には学習に用いられる教師なしデータの割合もあり、これが一因ではないかと考えられる。教師なしデータセットとして用いた Facescrub は有名人を撮影した画像データにより構成されており、主に正面方向から撮影されたデータが他の方向に比べて多い。ここで方向別の ROC-AUC の値で比較した場合、正面から撮影された場合の識別性能はあまり大きくないが、 $\pm 30^\circ$  から撮影された場合の識別性能に大きく差が生じている。これは教師なしデータの分布にしたがって正面に対する識別性能が強くなり、対して左右からの識別性能は下がってしまった結果と見ることができる。その結果、顔方向が当分のテストデータに対しては全体に対する識別性能が下がってしまったのではないかと考える。

また各撮影方向での識別精度についてもここで検討を行いたい。図 4.3 について正面方向から撮影した場合、いずれの手法でも ROC-AUC の値は 0.95 を超えており識別器の性能は非常に高い。対して  $\pm 30^\circ$  から撮影した場合についてはいずれの手法も正面と比較すると精度が下がった。これらは両目画像のみを入力としているおり、それが撮影方向の判別が難しいことが影響したという点が原因として考えられる。精度向上には顔方向といった撮影方向を判断できる情報を付与することが有効であると考えられる。またこの顔方向情報は、顔パーツの隠蔽に対して頑健に取得できることが望ましくこの手法の開発がアイコンタクト検出において今後の課題となる。

## 5. まとめ、今後の課題

本研究ではアイコンタクト検出器をより少ない教師ありデータセットにより作成するために、学習に半教師あり学習を用いる手法を提案した。半教師あり学習には Improved-GAN の枠組みを用い、GAN の学習を行う中でアイコンタクト識別器を生成する手法をとった。

実験では従来手法と提案手法の精度の比較及び半教師あり学習に用いるデータ数を変化させて比較検討を行った。結果として精度について、従来手法を上回ることができなかったが、提案手法において少ないデータ数で正答率 80% の精度の実現に成功したことから、アイコンタクト識別において半教師あり学習が有効であることが確認できたと考える。

今後の課題としては、半教師あり学習として従来手法の精度を実現するために他の GAN を用いた手法の導入、及びアイコンタクト識別器のさらなる精度向上のために顔方向を入力として扱う手法を立案する必要がある。

謝辞 本研究は科研費 17H01779, 26249029, 15H02738, および、JST, CREST, JPMJCR17A5 の支援を受けたものである。

## 参考文献

- [1] Jones, W. and Klin, A.: NIH Public Access, Vol. 504, No. 7480, pp. 427–431 (online), DOI: 10.1038/nature12715.Attention (2014).
- [2] イヴジネスト, ロゼットマレスコッティ, ジェロームベリシエ, 本田美和子, 辻谷真一郎: Humanitude(ユマニチュード)「老いと介護の画期的な書」, 株式会社トライアリスト東京 (2014).
- [3] 沖野祐介, 中澤篤志, 本田美和子, 石川翔吾, 竹林洋一, 西田豊明: 頭部装着型カメラを用いた介護スキル評価 (医用画像), 電子情報通信学会技術研究報告 = IEICE technical report: 信学技報, Vol. 116, No. 39, pp. 95–100 (2016).
- [4] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Networks, pp. 1–9 (online), DOI: 10.1001/jamainternmed.2016.8245 (2014).
- [5] Smith, B. A., Yin, Q., Feiner, S. K. and Nayar, S. K.: Gaze Locking: Passive Eye Contact Detection for Human-Object Interaction, *ACM Symposium on User Interface Software and Technology*, pp. 271–280 (online), DOI: 10.1145/2501988.2501994 (2013).
- [6] Ye, Z., Li, Y., Liu, Y., Bridges, C., Rozga, A. and Rehg, J. M.: Detecting bids for eye contact using a wearable camera, *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2015*, (online), DOI: 10.1109/FG.2015.7163095 (2015).
- [7] Yu Mitsuzumi, Atsushi Nakazawa, T. N.: DEEP eye contact detector: Robust eye contact bid detection using convolutional neural network, *British Machine Vision Conference* (2017).
- [8] Zhang, X., Sugano, Y. and Bulling, A.: Everyday Eye Contact Detection Using Unsupervised Gaze Target Discovery, *Proc. of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 193–203 (online), DOI: 10.1145/3126594.3126614 (2017).
- [9] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A.: It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2017).
- [10] Baltrusaitis, T., Robinson, P. and Morency, L. P.: OpenFace: An open source facial behavior analysis toolkit, *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, (online), DOI: 10.1109/WACV.2016.7477553 (2016).
- [11] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X.: Improved Techniques for Training GANs, pp. 1–10 (online), DOI: arXiv:1504.01391 (2016).
- [12] Radford, A., Metz, L. and Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, pp. 1–16 (online), DOI: 10.1051/0004-6361/201527329 (2015).
- [13] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [14] Ng, H. W. and Winkler, S.: A data-driven approach to cleaning large face datasets, *2014 IEEE International Conference on Image Processing, ICIP 2014*, pp. 343–347 (online), DOI: 10.1109/ICIP.2014.7025068 (2014).
- [15] Dai, Z., Yang, Z., Yang, F., Cohen, W. W. and Salakhutdinov, R.: Good Semi-supervised Learning that Requires a Bad GAN, (online), available from <http://arxiv.org/abs/1705.09783> (2017).
- [16] Li, C., Xu, K., Zhu, J. and Zhang, B.: Triple Genera-

tive Adversarial Nets, No. Nips (online), available from  
(<http://arxiv.org/abs/1703.02291>) (2017).