Learning Semi-Supervised Anonymized Representations by Mutual Information

Clément Feutry, Pablo Piantanida, Pierre Duhamel

LEARNING SEMI-SUPERVISED ANONYMIZED REPRESENTATIONS BY MUTUAL INFORMATION

C. Feutry *P. Piantanida* *P. Duhamel*

Laboratoire des Signaux et Systèmes
CentraleSupélec CNRS Université Paris-Saclay, Gif-sur-Yvette, France

ABSTRACT

This paper addresses the problem of removing from a set of data (here images) a given private information, while still allowing other utilities on the processed data. This is obtained by training concurrently a GAN-like discriminator and an autoencoder. The optimization of the resulting structure involves a novel surrogate of the misclassification probability of the information to remove. Several examples are given, demonstrating that a good level of privacy can be obtained on images at the cost of the introduction of very small artifacts.

1. INTRODUCTION

The problem we address, i.e. removing private information from a set of images, while still allowing any other search in the set is a problem that is supervised with respect to the private label, but unsupervised with respect to the original image. Hence the name of semi-supervised anonymization.

Related work. Some previous works addressed similar problems, very often in a more constrain setup. Reference [1] proposes a tool that allows to perform some emotion preservation while transferring identities using a variational generative adversarial network. This work is fully supervised contrary to our work, that only use private labels during training. Several other works, such as [2] or [3] address the problem of hiding tags (text label, QR code,...) in images. This problem is somewhat different from ours, because their target is to hide some regions of the image while we focus on hiding inherent private data. Another very related approach is image to image translation, such as [4, 5]. Instead of transferring the images to another domain, we provide a single "identity domain" representation where it is as hard as possible to track the initial identity (assuming that the private information is the identity).

De-identification is similar to anonymization but consists in the process of preventing someone personal ID from being revealed. The main difference between this concept and data anonymization is that some identifying information can be preserved in order to be relinked only by a trusted party or by the original data operator, whereas in the case of anonymization, no re-identification should be possible (e.g., [6, 7]).

Contributions. This paper presents a semi-supervised framework for removing a private information from image datasets, while keeping the distortion on the resulting images as small as possible. This model was used on two image datasets: a dataset of handwritten digits and a dataset of facial expression. It provided interesting results in the context of anonymization without a target task. Two cases should be distinguished: the case where the targeted private features are strongly connected to the other constitutive features of the image (denoted as intertwined) or not. Obviously, the later case is less difficult. Both cases are addressed below.

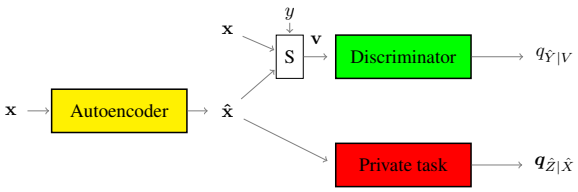

Fig. 1: Architecture (S is the selector block).

2. PRESENTATION OF THE PROBLEM

Since we wish to be able to perform the same classifications (except the private one) on the processed images and on the original images, the metric used to evaluate the output must be a mix between a fidelity metric (to ensure as much information as possible remains) and a private metric. However, the similarity metric cannot be a fixed metric (e.g., a traditional distortion). Indeed, a metric that performs well on a kind of data may not be as efficient on another data. Preliminary tests using fixed metrics yielded poor results.

This directed us to use a GAN-like discriminator as the flexible metric that is trained concurrently with an autoencoder. This has the advantage of keeping a wide range of applications and yet to obtain accurate and neat results. This flexibility is also useful to train an anonymizing autoencoder where images will need to change significantly to remove private features. The resulting architecture is displayed on Fig. 1. Random variables are in capital letters; $\widehat{\cdot}$ means computed over empirical distribution or estimate when over respectively a function or a random variable. Let $\mathcal{T}_n$ be a training set of

size n, where each element $(\boldsymbol{x}_i, z_i)$ is composed of $\boldsymbol{x}_i \in \mathcal{X}$, a real vector and of $z_i \in \mathcal{Z}$, the private label of this sample.

Our goal is to learn the parameters of the three networks: **(i)** $\boldsymbol{q}_{\hat{X}|X}$ the deep autoencoder; **(ii)** $q_{\hat{Y}|V}$ the discriminator; and **(iii)** $\boldsymbol{q}_{\hat{Z}|\hat{X}}$ the private classifiers. The selector S is before the discriminator and it allows a selection between $\mathbf{x}$ and $\hat{\mathbf{x}}$. The input $\mathbf{v}$ of the discriminator is the output of S which is driven by a variable Y. If $y = 1$ then $\mathbf{v} = \mathbf{x}$ otherwise $\mathbf{v} = \hat{\mathbf{x}}$.

Learning with anonymization constraints. *Consider the following constrained classification problem:*

$$\min_{(q_{\hat{X}|X}, q_{\hat{Y}|V}) \in \mathcal{F}} \Big\{ P_e\big(q_{\hat{X}|X}, q_{\hat{Y}|V}\big) \; : \; \min_{q_{\hat{Z}|\hat{X}} \colon \hat{\mathcal{X}} \to \mathcal{P}(\mathcal{Z})} P_e\big(q_{\hat{X}|X}, q_{\hat{Z}|\hat{X}}\big) \quad \geq 1 - \varepsilon \Big\}, \tag{1}$$

for a prescribed probability $1/|\mathcal{Z}| \leq \varepsilon < 1$, where the minimization is over the set of restricted encoders and classifiers $(q_{\hat{X}|X}, q_{\hat{Y}|V}) \in \mathcal{F}$ according to a model class $\mathcal{F}$.

The above expression requires representations with $(1 - \varepsilon)$-approximate guarantees (over all possible classifiers) w.r.t. the misclassification probability of the private labels. Obviously, ε can be replaced by a suitable positive multiplier $\lambda \equiv \lambda(\varepsilon)$ yielding a relaxed version of the objective.

$$\min \Big\{ P_e\big(q_{\hat{X}|X}, q_{\hat{Y}|V}\big) - \lambda \cdot P_e\big(q_{\hat{X}|X}, q^{\star}_{\hat{Z}|\hat{X}}\big) \Big\}, \tag{2}$$

where $q^{\star}_{\hat{Z}|\hat{X}}$ is the minimizer of $P_e\big(q_{\hat{X}|X}, q_{\hat{Z}|\hat{X}}\big)$. Expression (2) does not lead to a tractable objective for training $(q_{\hat{X}|X}, q_{\hat{Y}|V})$. Yet, it suggests a competitive game between two players: an adversary trying to infer the private labels Z from our representations $\hat{\boldsymbol{x}}$, by minimizing $P_e\big(q_{\hat{X}|X}, q_{\hat{Z}|\hat{X}}\big)$ over all possible $q_{\hat{Z}|\hat{X}}$, and a fidelity learner predicting the labels Y (i.e. the representations' fidelity), by optimizing a classifier $q_{\hat{Y}|V}$ over a prescribed model class $\mathcal{F}$. We can trade-off these two quantities via the autoencoder model $q_{\hat{X}|X}$. This idea will be further developed in the next section.

Cross-entropy loss*: Given two distributions $q_{\hat{X}|X} : \mathcal{X} \to \mathcal{P}(\mathcal{X})$ and $q_{\hat{Y}|V} : \mathcal{V} \to \mathcal{P}(\mathcal{Y})$, define the average (over representations) cross-entropy loss as:*

$$\ell\big(q_{\hat{X}|X}(\cdot|x), q_{\hat{Y}|V}(y|\cdot)\big) \coloneqq \big\langle q_{\hat{X}|X}(\cdot|x), -\log q_{\hat{Y}|V}(y|\cdot) \big\rangle \tag{3}$$

As usual, we shall measure the expected performance of $(q_{\hat{X}|X}, q_{\hat{Y}|V})$ via the risk*:*

$$\mathcal{L}(q_{\hat{Y}|V}, q_{\hat{X}|X}) \coloneqq \mathbb{E}_{p_{XY}}\big[\ell\big(q_{\hat{X}|X}(\cdot|X), q_{\hat{Y}|V}(Y|\cdot)\big)\big]. \tag{4}$$

It is not difficult to show from Fano's inequality that the lower bound on the misclassification task of the private labels is a monotonically decreasing function of the mutual information $\mathcal{I}(Z; \hat{X})$. This implies that any limitation of the mutual information between private labels Z and representations $\hat{X}$ will bound from below the probability of misclassification of private labels, whichever the chosen classifier $q_{\hat{Z}|\hat{X}}$. Besides, it is well-known that the cross-entropy provides a surrogate to optimize the misclassification probability of $\hat{Y}$, which motivates the cross-entropy loss. These information-theoretic bounds provide a mathematical objective in order to browse the trade-off (1) between all feasible misclassification probabilities $P_e\big(q_{\hat{X}|X}, q_{\hat{Y}|V}\big)$ as a function of the prescribed $(1-\varepsilon)$ probability. Therefore, the learner's goal is to select an autoencoder $q_{\hat{X}|X}$ and a classifier $q_{\hat{Y}|V}$ by minimizing jointly the risk and the mutual information. Yet, since p_{XYZ} is unknown one cannot directly measure neither the risk nor the mutual information. It is usual to measure the agreement of a pair of candidates using a training dataset whose empirical distribution $\hat{p}_{XYZ}$ is known. This yields an information-theoretic objective, being a surrogate of expression of eq. (2):

$$\min \big\{ \mathcal{L}_{\text{emp}}(q_{\hat{Y}|V}, q_{\hat{X}|X}) + \lambda \cdot \widehat{\mathcal{I}}(Z; \hat{X}) \big\}, \tag{5}$$

for a suitable multiplier $\lambda \geq 0$, where $\mathcal{L}_{\text{emp}}(q_{\hat{Y}|V}, q_{\hat{X}|X})$ denotes the *empirical risk* of eq (3) (i.e. averaged w.r.t. $\hat{p}_{XY}$). The mutual information must be empirically evaluated using $\hat{q}_{Z|\hat{X}}$ as being the posterior according to $q_{\hat{X}|X}\hat{p}_{XZ}$. Eq. (5) may be independently motivated by a different problem studying distortion-equivocation trade-offs [8].

Reconstruction learning with anonymization. The initial experiments performed with a similar training objective as the one introduced by [9] led to an unstable training and a poor trade-off between the degree of anonymity and the quality of the representation. This guides us to use instead a new adversarial training objective given below.

An examination of expression (5) shows that it cannot be optimized since the posterior distribution $\hat{q}_{Z|\hat{X}}$ is not computable in high dimensions. We will looser this surrogate by upper bounding the empirical mutual information $\widehat{\mathcal{I}}(Z; \hat{X}) = \widehat{\mathcal{H}}(Z) - \widehat{\mathcal{H}}(Z|\hat{X})$. The *empirical entropy* of Z can be upper bounded as follows:

$$\begin{aligned}
\widehat{\mathcal{H}}(Z) &\leq \mathbb{E}_{\hat{p}_Z}\big[-\log \hat{q}_{\hat{Z}}(Z)\big] && (6)\\
&\leq \mathbb{E}_{\hat{p}_Z}\mathbb{E}_{\hat{q}_{\hat{X}}}\big[-\log q_{\hat{Z}|\hat{X}}(Z|\hat{X})\big] && (7)\\
&\equiv \mathbb{E}_{\hat{p}_Z}\mathbb{E}_{\hat{p}_X}\big[\ell\big(q_{\hat{X}|X}(\cdot|X), q_{\hat{Z}|\hat{X}}(Z|\cdot)\big)\big] && (8)\\
&\coloneqq \mathcal{L}^{\text{obj}}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X}), && (9)
\end{aligned}$$

where (6) follows since the relative entropy is non-negative; (7) follows by the convexity of $t \mapsto -\log(t)$ and (8) follows from the definition of the cross-entropy loss. We will also resort to an approximation of the conditional entropy $\widehat{\mathcal{H}}(Z|\hat{X})$ by learning an adequate empirical cross-entropy risk:

$$\begin{aligned}
\widehat{\mathcal{H}}(Z|\hat{X}) &\approx \mathbb{E}_{\hat{p}_{XZ}}\big[\ell\big(q_{\hat{X}|X}(\cdot|X), q_{\hat{Z}|\hat{X}}(Z|\cdot)\big)\big]\\
&\equiv \mathcal{L}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X}),
\end{aligned} \tag{10}$$

which assumes a well-selected classifier $q_{\hat{Z}|\hat{X}}$, i.e., the resulting approximation error given by $\mathcal{D}\big(\hat{q}_{Z|\hat{X}} \| q_{\hat{Z}|\hat{X}} | \hat{q}_{\hat{X}}\big)$ w.r.t. the exact $q_{\hat{Z}|\hat{X}}$ is small enough. By combining expressions (9)

and (10), and taking the absolute value, we obtain :

$$\widehat{\mathcal{I}}(Z;\hat{X}) \precsim \left|\mathcal{L}^{\text{obj}}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X}) - \mathcal{L}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X})\right|, \quad (11)$$

that leads to our tractable learning objective, which is an approximation of expression (5), being the surrogate of (2):

$$\begin{aligned}\mathcal{L}_\lambda(q_{\hat{Y}|V}, q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X}) &\coloneqq \mathcal{L}_{\text{emp}}(q_{\hat{Y}|V}, q_{\hat{X}|X}) \\ &+ \lambda \cdot \left|\mathcal{L}^{\text{obj}}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X}) - \mathcal{L}_{\text{emp}}(q_{\hat{Z}|\hat{X}}, q_{\hat{X}|X})\right|, \quad (12)\end{aligned}$$

for a suitable classifier $q_{\hat{Z}|\hat{X}}$ and multiplier $\lambda \geq 0$ which tunes the trade-off between the discriminator and the private task.

The data representations provided by the autoencoder $\boldsymbol{q}_{\hat{X}|X}$ must blur the private labels Z features from the raw data X while preserving original data's relevant features. Note that (9) corresponds to the loss of a 'random guessing' classifier in which the representations $\hat{X}$ are independent of private labels Z. Thus, training encoders $q_{\hat{X}|X}$ to minimize (12) enforces the best classifier $q_{\hat{Z}|\hat{X}}$ to get closer – in terms of loss – to the random guessing classifier.

Training objectives. The terms of loss function (12) are successively the binary cross-entropy of discriminator, the cross-entropy of the private branch and the objective cross-entropy which expressions follow:

$$\begin{aligned}\mathcal{L}_{\text{emp}}(q_{\hat{Y}|V}, \boldsymbol{q}_{\hat{X}|X}) = -\frac{1}{n}\sum_{i=1}^{n}\Big[&y_i \log q_{\hat{Y}|V}(\cdot|\boldsymbol{v}_i) \\ &+(1-y_i)\log(1-q_{\hat{Y}|V}(\cdot|\boldsymbol{v}_i))\Big], \quad (13)\end{aligned}$$

$$\mathcal{L}_{\text{emp}}(\boldsymbol{q}_{\hat{Z}|\hat{X}}, \boldsymbol{q}_{\hat{X}|X}) = \frac{1}{n}\sum_{i=1}^{n}\left\langle \boldsymbol{e}(z_i), -\log \boldsymbol{q}_{\hat{Z}|\hat{X}}(\cdot|\boldsymbol{x}_i)\right\rangle, \quad (14)$$

$$\mathcal{L}^{\text{obj}}_{\text{emp}}(\boldsymbol{q}_{\hat{Z}|\hat{X}}, \boldsymbol{q}_{\hat{X}|X}) = \frac{1}{n}\sum_{i=1}^{n}\left\langle \hat{\boldsymbol{p}}_Z, -\log \boldsymbol{q}_{\hat{Z}|\hat{X}}(\cdot|\boldsymbol{x}_i)\right\rangle. \quad (15)$$

$\boldsymbol{e}(z_i)$ are "one-hot" vectors (z_i component is 1 and the others 0) of the true labels of sample $i = [1:n]$. $\hat{\boldsymbol{p}}_Z$ is the empirical distribution of the private labels.

Each branch, i.e., $q_{\hat{Y}|V}$ and $\boldsymbol{q}_{\hat{Z}|\hat{X}}$, is trained to minimize the associated cross-entropy loss (respectively (13) and (14)), whereas the autoencoder $q_{\hat{X}|X}$ is trained to minimize eq. (12) (minimizing the cross-entropy loss with respect to $\hat{Y}$ predictor while maximizing the new adversarial loss defined with respect to the $\widehat{Z}$ predictor).

Training procedure. First the autoencoder is pre-trained using a pixel to pixel metric loss for a few epochs. Then the discriminator is pre-trained by alternating between: **(i)** updates on the discriminator parameters while autoencoder's parameters are frozen, with inputs going through the selector being a combination of autoencoded and original samples; **(ii)** updates on the autoencoder while discriminator's parameters are frozen. This allow the discriminator to learn the difference between both distributions. Finally, updates between the autoencoder on one side and the discriminator and the private branch on the other side are done alternatively. It is during this step that the privacy/fidelity trade-off occurs in the autoencoder.

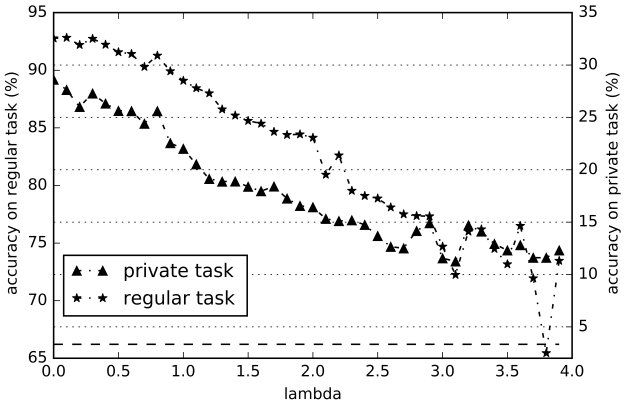

Fig. 2: Accuracy as a function of $\lambda \in [0, 4.0]$ on Pen-digits dataset. Horizontal black dashed line is the random guessing classifier over the user-ID (3.33%).

3. SIMULATION RESULTS

The datasets used are doubly labelled: one is the private task classification and the other is used for the regular task classification in the performances assessment.

The **Pen-digits** dataset, from [10] provides the coordinates of digitally acquired pen movements of 44 persons (30 are in the training set and 14 in the test-set) writing digits from 0 to 9. We only used the training set (only 30 identities) which was randomly split into training, validation and test sets (size 5494, 1000 and 1000, respectively), sharing images of the same 30 persons. After being drawn, each image was down-sampled into a 20x20 grey-scale image.

The **FERG** (Facial Expression Research Group) dataset [11] contains 55767 annotated face synthetic images of six stylized characters modeled with the MAYA software. These images depict the seven following facial expressions: 'neutral', 'anger', 'fear', 'surprise', 'sadness', 'joy' and 'disgust'. Original 256x256 colour images have been pre-processed into re-sized 8-bit grey-scale 50x50 images.

Pen-digits results The private task labels in this dataset correspond to the writer identity. Hence, we want to transform handwritten digit images into other handwritten digit images while removing identity features and keeping relevant features about the original image digit.

Once the model is trained, original images are autoencoded into transformed images. Assessing the performance in a fair and automated manner is hard for it implies comparing and judging pictures. We processed images of the whole dataset and associated them with their original labels. Two as-

targeted identity	0	2	3	4	5
identity accuracy (%)	42.47	38.25	36.21	32.77	31.62
emotions accuracy (%)	72.45	77.46	72.32	66.79	72.95

Table 1: Accuracies for anonymization (supervised during training) with different targeted IDs and with the accuracies of emotions recognition (unsupervised during training) to measure conservation of unlabeled intertwined features. These are obtained by training a classifier on the processed dataset.

sessments networks are then trained on the processed images, trying to recover both labels (ID and digit). The evaluation, plotted on Fig. 2, gives the accuracy on both tasks.

This assessment method judges a semi-supervised transformation in a fully supervised manner, yet it is a good compromise to quantify the effect on the ID features and on some other features that we wanted to keep. This testing is realistic since we used all the available ID information to train a ID features extractor from the transformed images. It is the best case scenario for an attacker wanting to recover the ID from the processed dataset. Note that even if testing on only one non-private feature is not ideal, the fidelity assessment is performed on a feature not used during the training of the model.

FERG dataset. The private features to remove from the representations are the users identity labels.

Intertwined task method. The standard procedure explained before failed on the FERG dataset. Despite producing images that fool human eyes, they fail to fool a neural network. This is caused by the intertwined nature of the data. Identity traits and emotions are quite mixed. Our previous method led to a network that can produce visually anonymized results, but not a computer validated anonymization. To improve the anonymization on this kind of datasets, we derived a second method which specificity is restricting to one chosen identity the discriminator's train samples. This method although providing a small gain in anonymization was not sufficient. To perform a good anonymization we provided a protocol of substitution. Once all images are transformed into the chosen identity, each is compared to the chosen identity's real samples. Using pixel to pixel euclidean distance images are substitute by the closest chosen identity's training set image. The substitution is made really easy because the encoder shifts the representation of the original images very close to the selected identity while maintaining the underlying structure of the emotion features.

FERG Results. As stated above, two methods are associated with this particular framework. The first method restricts the discriminator to only one chosen ID and produces qualitative results shown on the Figs. 3 and 4. Whereas the substitution method produces measurable results shown in Table 1.

Remark. One could argue that even if the dataset is anonymized, the chosen identity is still disclose because every sample wears the face of one preexisting identity. This is not a problem at all considering one can add an artificially identity in the databse to be the template of the anonymized faces, while still using our framework. Indeed the fact that the intertwined method allows to choose the final identity from the training set make it possible.

Fig. 3: Evolution of the reconstructed images for a training and a validation samples (respectively top and bottom line). The resulting identity is fixed by the operator. The first image is the end of the pre-training phase: the autoencoder is trained to reproduce its entry (due to the pre-training's performance, this image is identical to the input sample). Going right follows the increase of the training epochs one by one. The most impressive changes occur over these first epochs. The following epochs (not show here) only refine the details of the encoded image to make it look genuine, resulting in the far right images that are the end training results.

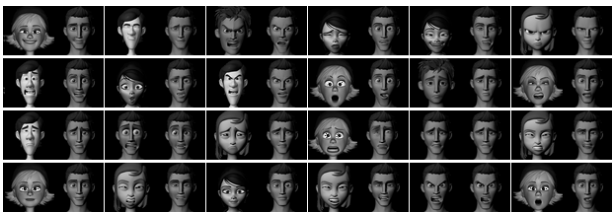

Fig. 4: 24 couples of images composed of the sample and corresponding model output. All are randomly taken from the test set. Some representations are showing artifacts (missing an eye), e.g., representation at positions (line,column): (3,4),(3,2). Other shift emotions (1,4).

4. CONCLUSION

We succeed in providing an anonymization framework where only the private labels are used to train the model. This is possible via the introduction of a new loss (expression (12)), an adapted architecture and training protocol. To train this loss, we took advantage of and reused widely known and efficient methods such as GAN discriminator and autoencoder, and we managed to perform a smart combination of both to design a working architecture that was suited to our loss.

The results of this paper are made even more impressive because of the visual representation. The network we designed manages to affect the hidden representation of the data in the autoencoder to shift private bit of information toward a dataset universal template with minimum cost to the other parts of the information. The generalized information that we can deduce is that this method allows a structure adapted spatial shift in the learnt features space: shifting from identity to another without major loss of sub-cluster structure (here emotions or digits). The usage of the learnt discriminator allows a neat output image, pixel perfect in most cases.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 797805.

5. REFERENCES

[1] Jiawei Chen, Janusz Konrad, and Prakash Ishwar, "Vgan-based image representation learning for privacy-preserving facial expression recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[2] Harrison Edwards and Amos Storkey, "Censoring representations with an adversary," *arXiv preprint arXiv:1511.05897*, 2015.

[3] N. Raval, A. Machanavajjhala, and L. P. Cox, "Protecting visual secrets using adversarial nets," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 1329–1332.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017.

[5] Ming-Yu Liu, Thomas Breuel, and Jan Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., pp. 700–708. Curran Associates, Inc., 2017.

[6] Blaz Meden, Refik Can Malli, Sebastjan Fabijan, Hazim Kemal Ekenel, Vitomir Struc, and Peter Peer, "Face deidentification with generative deep neural networks," *CoRR*, vol. abs/1707.09376, 2017.

[7] Yuezun Li and Siwei Lyu, "De-identification without losing faces," *CoRR*, vol. abs/1902.04202, 2019.

[8] Joffrey Villard and Pablo Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *IEEE Trans. Inf. Theor.*, vol. 59, no. 6, pp. 3668–3692, June 2013.

[9] Yaroslav Ganin and Victor S. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 2015, pp. 1180–1189.

[10] F Alimoglu and E Alpaydin, "Methods of combining multiple classifiers based on different representations for pen-based handwriting recognition," in *Proceedings of the Fifth Turkish Artificial Intelligence and Artificial Neural Networks Symposium (TAINN 96)*, Jun 1996.

[11] Deepali Aneja, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones, "Modeling stylized character expressions via deep learning," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 136–153.