様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 ２４ 日現在

機関番号：６２６１５
研究種目：研究活動スタート支援
研究期間：2020～2021
課題番号：２０Ｋ２３３５５

研究課題名（和文）Facial Privacy and Forensic in The Wild: Explainable End-to-End Networks for Multi-Face Anonymization and Multi-Face Forgery Detection

研究課題名（英文）Facial Privacy and Forensic in The Wild: Explainable End-to-End Networks for Multi-Face Anonymization and Multi-Face Forgery Detection

研究代表者

レ　チュンギア（Le, Trung-Nghia）

国立情報学研究所・情報社会相関研究系・特任研究員

研究者番号：００８８４４０４

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：偽造データの合成コストを削減するために、偽造ワークフローを開発しました。 私たちのフレームワークは、ディープフェイクジェネレーターを繰り返しトレーニングすることなく、非ターゲットフェイススワッピング用のGANモデルを使用して、無限の数の偽の個人IDを生成できます。 このフレームワークは、ディープフェイクの生成と匿名化に大きな可能性を秘めています。 また、多面的な偽造の検出と実際のセグメンテーションのために、高品質の画像を使用して新しい大規模なデータセットを作成しました。 また、これらのタスクの評価と進行を容易にするためのベンチマークスイートも紹介しました。

研究成果の学術的意義や社会的意義
We published a book chapter to introduce general knowledge about deepfake for beginners and/or students. We expect that our book chapter is helpful for beginners to understand deepfake and use these techniques correctly.

研究成果の概要（英文）：We developed a forgery workflow to reduce the cost of synthesizing fake data. Our framework can generate an infinite number of fake individual identities using GAN models for non-target face-swapping without repeatedly training a deepfake generator. This framework has great potential in deepfake generation and face anonymization. We also created a new large-scale dataset with high-quality images for multi-face forgery detection and segmentation in-the-wild. It consists of 115K unrestricted images with 334K human faces. We also presented a benchmark suite to facilitate the evaluation and advancement of these tasks. Our work was published at ICCV 2021, a top-tier conference in computer vision.

研究分野：Applied Computer VIsion

キーワード：deepfake generation　deepfake detection　deepfake segmentation

１．研究開始当初の背景

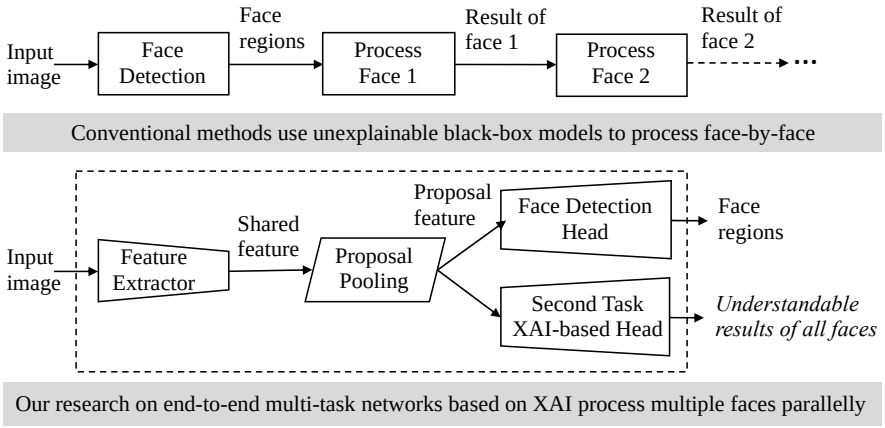
By learning billions of images/videos uploaded and shared on social networks, recognition-based automatic systems (i.e., auto-tagging, or auto-captioning) help our life more convenient but also raise a concern about privacy. To protect privacy, it is essential to develop methods to de-identify human faces whose generated images/videos should be realistic and fitting the given situation but still have the ability of defense from face recognition. Meanwhile, falsified images/videos with a realistic appearance synthesized by deep learning (i.e., DeepFakes, Face2Face) raise significant concerns about individual harassment or deceptive criminals. To address threats imposed by spoofing/impersonating attacks, it is essential for the development of countermeasures to identify facial forgeries in digital media. These two research fields have a great relationship and complement each other. Hence, understanding face anonymization can help to improve facial forgery detection and vice versa.

Conventional methods of face anonymization and facial forgery detection need given face regions as the input. Therefore, these methods can process only a face at the time, and their performance heavily depends on the accuracy of face detection methods. Meanwhile, it takes so much time to process multiple faces sequentially. Hence, it is essential for the development of effective methods to process multiple faces at the same time from an input image. In addition, conventional methods were evaluated on only laboratory environments with a simple background and single clear front face. Hence, these methods have not been ready to be deployed in the real world, whose contexts are much more diverse and challenging than the simple staged scenarios.

Most of conventional models were developed as a poorly understood black-box, in which human totally replied on the models without understanding of their behaviors. Specially, people are difficult to identify fake faces, which recently have been synthesized with extremely realistic appearance so that they accept all outputs of models regardless their reliability. Blindly accepting unexplainable outcomes is dangerous because the models are not infallible and may be vulnerable to adversarial attacks. To address these issues, eXplainable AI (XAI) aims to develop explainable and interpretable solutions such that results can be understood by human experts. For instance, besides mapping faces to real/fake labels, XAI can additionally identify manipulated areas to improve the explainability and convincibility of results. Therefore, XAI is essential for the development of trusted AI systems on the demands of society, in which people can trust outcomes of the AI systems to inform their decisions.

２．研究の目的

The main objective of this research is to explore multi-face anonymization/forgery detection in the wild. It can overcome limitations of conventional methods as we mentioned above.



In particular, we investigate two key scientific questions: 1) How can we combine knowledge of biometric and computer vision for the construction of end-to-end multi-task networks, which can perform multi-face anonymization/forgery detection in the wild at a real-time speed? 2) How can we develop trustworthy and robust models based on XAI?

３．研究の方法

We found that to increase reliableness of deep models based on XAI, we can perform segmentation to identify pixel-wise regions which are manipulated. Therefore, we investigated instance-level manipulated face detection and segmentation. In particular, we studied end-to-end multi-task networks to process multiple faces parallelly in short time. The network consists of a shared backbone to extract features, a proposal pooling component to pool features for all possible facial proposals, and head-networks corresponding with multiple tasks. The first head-network eliminates non-face regions and refines bounding box of face regions. The second head-network aims to cover all faces with anonymized features or categorize each face into real-face/fake-face classes; network of each task is developed using the complementary from the remained task. This head-network can also segment manipulated pixels in the fake-faces.

We also constructed new realistic datasets with multiple faces in natural scenes to evaluate the performance of multi-task methods. Differently from existing biometric datasets, which contains only a single clear font face with the simple background, constructed datasets can support multiple faces with various poses, emotions in natural scenes. In this way, we can develop computational models in both biometric and computer vision domains.

４．研究成果

We developed a forgery workflow to reduce the cost of synthesizing fake data. Our framework can generate an infinite number of fake individual identities using GAN models for non-target face-swapping without repeatedly training a deepfake generator. This framework has great potential in deepfake generation and face anonymization. We also created a new large-scale dataset with high-quality images for multi-face forgery detection and segmentation in-the-wild. It consists of 115K unrestricted images with 334K human faces. We also presented a benchmark suite to facilitate the evaluation and advancement of multi-face forgery detection and segmentation. Our work was published at ICCV 2021, a top-tier conference in computer vision.

We published a book chapter to introduce deepfake for beginners. Our chapter overviewed deepfake generation and detection methods from the viewpoint of technical evolution in computer vision. In particular, it described deepfake generation methods in different categories and analyzed their limitations. In addition, it clarified the different tasks of deepfake detection, from conventional classification to modern end-to-end detection and segmentation. It further discussed the limitations of deepfake detection methods and suggested solutions for improving the robustness of deepfake detection. Finally, it suggested a future direction for deepfake detection.

Peer-reviewed Book Chapters and Journals:

- Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, “Robust Deepfake On Unrestricted Media: Generation And Detection”, Frontiers in Fake Media Generation and Detection, 2022.
- Trung-Nghia Le, Tam V. Nguyen, Minh-Triet Tran, “Contextual Guided Segmentation Framework for Semi-supervised Video Instance Segmentation”, Journal of Machine Vision and Applications (MVA), 2022.
- Trung-Nghia Le, Yubo Cao, Tan-Cong Nguyen, Minh-Quan Le, Khanh-Duy Nguyen, Thanh-Toan Do, Minh-Triet Tran, Tam V. Nguyen, “Camouflaged Instance Segmentation In-The-Wild: Dataset, Method, and Benchmark Suite”, IEEE Transactions on Image Processing (T-IP), 2022.
- Vatsa S Patel, Zhongliang Nie, Trung-Nghia Le, Tam V. Nguyen, “Masked Face Analysis via Multi-task Deep Learning”, Journal of Imaging, 2021.
- Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, Tam V. Nguyen, “MirrorNet: Bio-Inspired Camouflaged Object Segmentation”, IEEE Access, 2021.

Peer-reviewed Conference and Workshop Papers:

- Minh-Quan Le, Trung-Nghia Le, Tam V. Nguyen, Isao Echizen, Minh-Triet Tran, “GUNNEL: Guided Mixup Augmentation and Multi-View Fusion for Aquatic Animal Segmentation”, CVPR Workshop on CV4Animal, 2022.
- Khanh-Duy Nguyen, Huy H. Nguyen, Trung-Nghia Le, Junichi Yamagishi, Isao Echizen, “Effectiveness of Detection-based and Regression-based Approaches for Estimating Mask-

Wearing Ratio", FG4COVID19 Workshop, 2021

- Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, "OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild", International Conference on Computer Vision (ICCV), 2021.
- Marc Treu, Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, "Fashion-Guided Adversarial Attack on Person Segmentation", CVPR Workshop on Media Forensics, 2021.
- Trung-Nghia Le, Tam V. Nguyen, Quoc-Cuong Tran, Lam Nguyen, Trung-Hieu Hoang, Minh-Quan Le, Minh-Triet Tran, "Interactive Video Object Mask Annotation", AAAI Conference on Artificial Intelligence, 2021.
- Trung-Nghia Le, Vuong Nguyen, Cong Le, Tan-Cong Nguyen, Minh-Triet Tran, Tam V. Nguyen, "CamouFinder: Finding Camouflaged Instances in Images", AAAI Conference on Artificial Intelligence, 2021.

Others:

- Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, "Deepfake 生成と検出の現状", インフォデミック時代の AI とサイバーセキュリティ, 2022.
- Futa Waseda, Sosuke Nishikawa, Trung-Nghia Le, Huy H. Nguyen, Isao Echizen, "Closer Look at the Transferability of Adversarial Examples: How They Fool Different Models Differently", arXiv:2112.14337, 2021.

５．主な発表論文等

〔雑誌論文〕　計11件（うち査読付論文　11件／うち国際共著　11件／うちオープンアクセス　10件）

１．著者名 Trung-Nghia Le, Vuong Nguyen, Cong Le, Tan-Cong Nguyen, Minh-Triet Tran, Tam V. Nguyen	４．巻 0
２．論文標題 CamouFinder: Finding Camouflaged Instances in Images	５．発行年 2021年
３．雑誌名 AAAI Conference on Artificial Intelligence	６．最初と最後の頁 16071-16074
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Trung-Nghia Le, Tam V. Nguyen, Quoc-Cuong Tran, Lam Nguyen, Trung-Hieu Hoang, Minh-Quan Le, Minh-Triet Tran	４．巻 0
２．論文標題 Interactive Video Object Mask Annotation	５．発行年 2021年
３．雑誌名 AAAI Conference on Artificial Intelligence	６．最初と最後の頁 16067-16070
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Marc Treu, Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen	４．巻 0
２．論文標題 Fashion-Guided Adversarial Attack on Person Segmentation	５．発行年 2021年
３．雑誌名 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops	６．最初と最後の頁 943-952
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CVPRW53098.2021.0010	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen	４．巻 0
２．論文標題 OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild	５．発行年 2021年
３．雑誌名 IEEE/CVF International Conference on Computer Vision	６．最初と最後の頁 10117-10127
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICCV48922.2021.00996	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Khanh-Duy Nguyen, Huy H. Nguyen, Trung-Nghia Le, Junichi Yamagishi, Isao Echizen	4．巻 0
2．論文標題 Effectiveness of Detection-based and Regression-based Approaches for Estimating Mask-Wearing Ratio	5．発行年 2021年
3．雑誌名 IEEE International Conference on Automatic Face and Gesture Recognition Workshops	6．最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/FG52635.2021.9667046	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Minh-Quan Le, Trung-Nghia Le, Tam V. Nguyen, Isao Echizen, Minh-Triet Tran	4．巻 0
2．論文標題 GUNNEL: Guided Mixup Augmentation and Multi-View Fusion for Aquatic Animal Segmentation	5．発行年 2022年
3．雑誌名 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops	6．最初と最後の頁 1-15
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, Tam V. Nguyen	4．巻 9
2．論文標題 MirrorNet: Bio-Inspired Camouflaged Object Segmentation	5．発行年 2021年
3．雑誌名 IEEE Access	6．最初と最後の頁 43290-43300
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2021.3064443	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Vatsa S Patel, Zhongliang Nie, Trung-Nghia Le, Tam V. Nguyen	4．巻 7
2．論文標題 Masked Face Analysis via Multi-task Deep Learning	5．発行年 2021年
3．雑誌名 Journal of Imaging	6．最初と最後の頁 1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/jimaging7100204	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Trung-Nghia Le, Yubo Cao, Tan-Cong Nguyen, Minh-Quan Le, Khanh-Duy Nguyen, Thanh-Toan Do, Minh-Triet Tran, Tam V. Nguyen	4．巻 31
2．論文標題 Camouflaged Instance Segmentation In-The-Wild: Dataset, Method, and Benchmark Suite	5．発行年 2022年
3．雑誌名 IEEE Transactions on Image Processing	6．最初と最後の頁 287-300
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TIP.2021.3130490	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Trung-Nghia Le, Tam V. Nguyen, Minh-Triet Tran	4．巻 24
2．論文標題 Contextual Guided Segmentation Framework for Semi-supervised Video Instance Segmentation	5．発行年 2022年
3．雑誌名 Machine Vision and Applications	6．最初と最後の頁 1-19
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00138-022-01278-x	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen	4．巻 0
2．論文標題 Robust Deepfake On Unrestricted Media: Generation And Detection	5．発行年 2022年
3．雑誌名 Frontiers in Fake Media Generation and Detection	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関