Tanmoy Mondal
Research Statement
February 18, 2024

Drawing from my prior research experiences and the various opportunities, I've encountered throughout my professional journey which sometimes required me to adjust the research topics with available resources and logistical feasibility (e.g. funding amount and duration, domain knowledge). Considering my domain expertise, personal interests, and the relevance of these topics in real-world applications, I've identified several areas that I find intriguing to explore further. Since completing my Ph.D., I've worked either as a post-doctoral researcher or in the industry as a scientist, where I have been primarily engaged in projects that are mainly assigned to me (which intrinsically built my knowledge-base on certain problem domains) rather than pursuing my research direction. However, given the chance to conduct independent research, I am enthusiastic about delving into the following topics.

I) Limiting Image Manipulation & Steganography : The proliferation of novel image editing techniques has greatly enriched our visual world. But manipulating images has never been easier, with new powerful editing tools appearing day by day. In the past, creating a multimedia disinformation campaign demanded advanced abilities, and attackers were limited to basic techniques like copying, replicating, or deleting elements within an image, commonly referred to as "cheap fakes". However, due to the rapid advancement of deep learning, image manipulation tools have become more accessible and potent. This enables users to effortlessly produce images of nonexistent individuals or create convincing deepfakes.

To conquer the image falsification, one of the notable way is *steganography*, which is defined as the art of hiding some secret data by embedding it into a host medium that is not secret (e.g. image). Accordingly, image *steganography* refers to the process of hiding data within an image file. The application areas are in digital communication, image copyright protection, information certification & authentication etc. The goal of good *steganography* is that the hidden data should not be detected under *steganalysis*, which is the countermeasure of steganography. Therefore, image steganography essentially asks for a powerful image representation mechanism that can effectively approximate the host image with the "noise" of the hidden images or messages. This process is also expected to be reversible because the hidden image or message should be recovered from the container image during the decoding process of image steganography.

I, therefore, seek to work on learning computer vision techniques that would be capable of detecting manual image manipulation (e.g. using software like Photo-shop to perform splicing, copy-move, removal, and image enhancement) in real-time, where the goal will be to create a system which performs well with the limited amount of training data. Furthermore, I want to work on the techniques to combat the generative models or diffusion model-based creation of realistic image editing and manipulation.

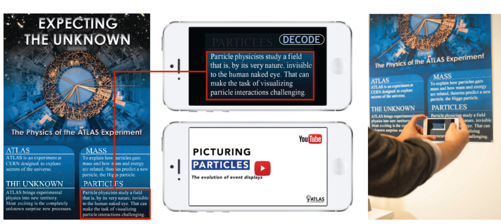

Figure 1: A bar-code pattern is encoded inside the text characters. The smart phone application allows the user to take a snapshot (right) and decode the hidden bar-code which is a Youtube link. Image credit: [1]

i) Forgery detection in administrative document images & natural images : Processes such as insurance claims, expense management or transmission of sensitive information (payslips, bills, etc) are some examples of the use of this technology on significant daily operations. Recently, we published [2] a technique to verify passports images. Continuing in the same direction, we are currently working on the techniques to authenticate receipts, vouchers, bills (telephone, water etc.). The objective is to develop a technique to detect and localize the frauds in real-time.

In the case of natural images, SOTA techniques are not yet well performing enough for in-the-wild deployment, mainly due to the deficiencies in these areas [4] : *a)* limited generalization (i.e. cope with out-of-distribution manipulations); *b)* limited robustness (i.e. robust to re-compression, re-sizing, etc.); *c)* insufficient detection performance. I would like to focus my work to handle these existing deficiencies.

Figure 2: A message (QR code) is encoded within an image or video then the image/video is displayed. The hidden message is recovered from the taken photograph. Image credit: [3]

Following the same path of research, I would also like to work on various techniques to *steganography* to make the host image more robust to various image deformations (e.g. geometric deformations like scaling, rotation, translation, etc.) and noise (e.g. image compression noise, print-scan noise, etc.). Furthermore, I am also interested in making image *steganography* applications more efficient in practice which will be capable to embed as much hidden data as possible into the host image. I would like to work on *steganography* techniques like the one, shown in Fig. 1, where the user-specified information is embedded in the text by perturbing the glyphs of text characters, while preserving the text content. Furthermore, the techniques to embed one image in another host image where it is important to take care the robustness, noise and compression interference is a challenging area to work on. For example, in Fig. 2, a QR-code image is encoded within an image.

II) Artificial Data Generation: The vision-based models are heavily reliant on annotated and curated data which is both costly and frequently unethical, often involving underpaid labor or the unauthorized scraping of data. Moreover, the emphasis on "big data" neglects rare or unconventional scenarios, undermining the robustness of these systems in real-world applications. I therefore seek to work on models, capable of automatic data (meaningful and useful) generation. In this direction of work, based on my previous experience and interest, I would like to work on the following three specific problem statements :

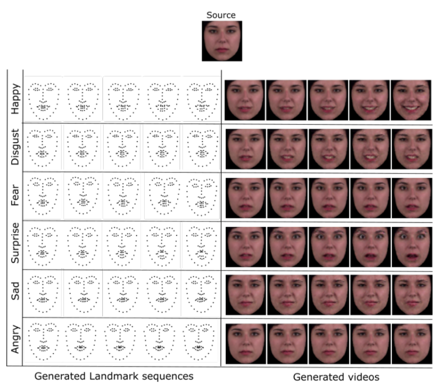

Figure 3: Given a neutral image, the algorithm should be able to generate sequences of facial landmarks for different facial expressions and transform them to videos. Image credit: [5]

i) Dynamic Facial Expression Generation : Facial expression holds significant importance in social interactions and spans various potential applications ranging from human-computer interaction to medical and psychological research. Creating authentic facial expressions holds numerous possibilities, such as enhancing facial expression classifiers through data augmentation, training robust models for facial recognition, animating robots and avatars, developing computer games, and aiding in facial surgery planning. While generative models have demonstrated remarkable performance in various image generation tasks, particularly in static facial expression synthesis, the exploration of dynamic facial expression synthesis remains relatively limited. Additionally, dynamic facial expressions are characterized by their temporal evolution, posing a more intricate challenge compared to static expressions. Thus, my focus will be on synthesizing realistic facial expression videos, acknowledging the complexity of capturing the dynamic progression of facial expressions.

ii) Dynamic Pose Generation of Person : In this sector, I am interested in generating images of non-rigid objects that often possess a large variation of deformation and articulation. More specifically, I am interested in transferring a person from one pose to another (as shown in Fig. 4). The topic of "pose generation" or "pose transfer" is valuable in many tasks such as video generation with the sequence of poses and data augmentation for person re-identification. The pose transfer can be exceptionally challenging, particularly when given only partial observations of the person. For example, when it is needed to infer the unobserved body parts to generate the target poses and views. Furthermore, complication appears when images captured from varying angles or poses can exhibit significant differences in appearance. Consequently, the generator must effectively encompass the extensive variations present within the distribution of images.

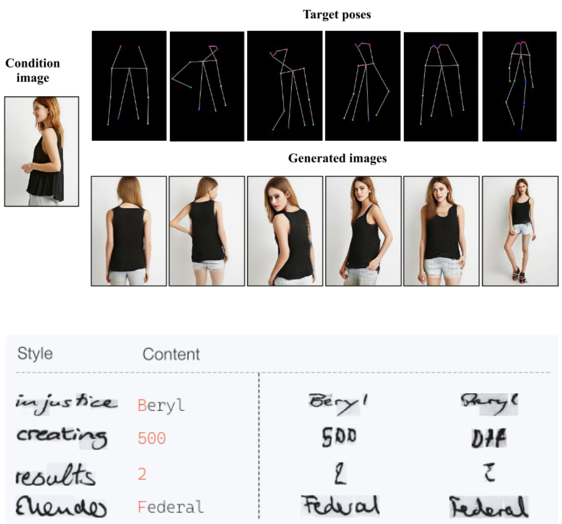

Figure 4: (Top:) The objective is to generate images based on different target poses. Image credit: [6]. (Bottom:) The word image based on given writing style. Image credit: [7]

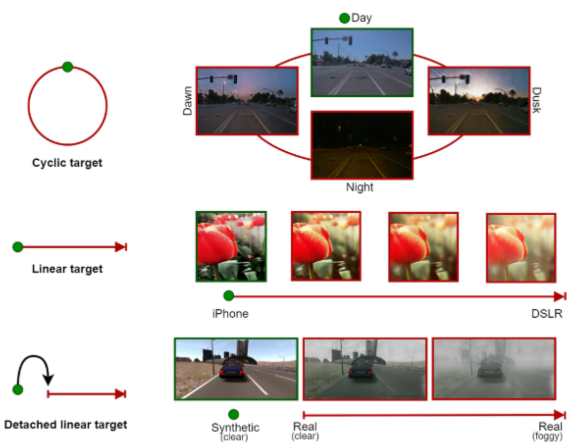

Figure 5: The i2i translation from day images, shallow depth of field from in-focus images and synthetic foggy images. Image credit: [7]

iii) Generation of Styled Handwritten Character & Word Images : Styled handwritten text generation is an emerging research area aimed at producing writer-specific hand-written text images by mimicking their calligraphic style (shown at the bottom of Fig. 4). The practical applications range from the synthesis of high-quality training data for personalized Hand Written Text Recognition (HTR) models to the automatic generation of handwritten notes for physically impaired people. I would like to work on this topic by focusing on the *Few-shot Styled offline* hand-written text generation tasks, in which we have just a few examples of images of the writer's style to mimic.

III) Physics-based Analysis & Generation of Images: The classical image-to-image (i2i) translation networks learn translations between domains, applying to the context of source images and a target appearance, learned from the dataset. There are several applications such as "neural photo editing", along with robotics-oriented tasks such as time-of-day or weather selection. Despite impressive leaps forward with the unpaired, multi-target, or continuous i2i, there are still important limitations. Specifically, to learn the complex continuous i2i translations, the existing works require supervision on intermediate domain points. Furthermore, most of the i2i techniques assume piece-wise or entire linearity of the domain translations (e.g. daytime) or continuous ones which are costly or impractical to label (e.g. fog, rain, etc.) (see Fig. 5). I would like to work on i2i techniques for learning non-linear continuous translations with unsupervised target data. The model would be trained by using physics-inspired models for guidance, while the relaxing model depends on continuous disentanglement of domain features. In our recent work in [8], we proposed a deep learning-based architecture to automatically simulate the underwater effects where only "dehazing" like physics based image formation equation is known to the network, and the additional degradation due to the other unknown factors is inferred in a data-driven way. The successful results of this work, opens-up the avenues to adapt and explore the similar techniques to work on physics model based "dehazing" network to remove "haze" from images [9]. Furthermore, thanks to our work in [8], where we have dealt with *depth image generation from RGB image* [10] as intermediate process, I would like to also explore the techniques, related to depth image estimation where the physics based image formation model can be used as a guiding factor for i2i translation.

V) Huge Time-Series Analysis : The discovery of time series motifs and anomalies has emerged as one of the most useful primitives in time series data mining. Researchers have shown its utility for exploratory data mining, summarization, visualization, segmentation, classification, clustering, and rule discovery. Although there has been more than a decade of extensive research, there is still a little amount of work, done in the direction to allow the discovery of time series motifs and anomalies in the presence of missing data, despite the well-documented ubiquity of missing data in scientific, industrial and medical datasets. Recently, the "Matrix Profile" has emerged as the state-of-the-art for finding time series motifs and anomalies in large datasets. Furthermore, we recently have shown that "k-nearest neighbors (kNN)" based matrix-profile [11] performs better than (in certain application domains) classical 1NN based matrix profile. In another recent work [1], we have shown that non-normalized euclidean distance based matrix profile performs better than classical z-normalization based matrix profile techniques.

But in general the matrix profile based techniques not only fails in the case of missing data but is also dependent on a user-given parameter i.e. length of time series motifs and anomalies which we expect or wish to find. In many cases, this is a reasonable limitation as the user may utilize out-of-band information or domain knowledge to set this parameter. However, in truly exploratory data mining, a poor choice of this parameter can result in failing to find unexpected and exploitable regularities in the data. Hence, I would like to work on motifs and anomaly discovery techniques which are parameter-free. Furthermore, the success of motifs or anomaly detection techniques is not well generalized to the multi-dimensional time series. The main problem appears to be that in any $N-$dimensional time series, the anomaly will generally only manifest itself on K dimensions only of the time series with $K < N$ (typically $K << N$), and the inevitable small amounts of noise on the remaining $N - K$ dimensions will tend to overshadow the signal provided by the anomalous time series. Hence, I am interested in working on the problem of motifs and anomaly detection of multi-dimensional time series.

References

[1] Chang Xiao, Cheng Zhang, and Changxi Zheng. FontCode: Embedding information in text documents using glyph perturbation. *ACM Transactions on Graphics*, 2018.

[2] Zuheng Ming Petra Gomez-Krämer Mickaël Coustaty Nicolas Sidere Jean-Christophe Burie Musab Al-Ghadi, Tanmoy Mondal. Identifying fraudulent identity documents by analyzing imprinted guilloche patterns. In *Springer Multimedia Tools and Applications Journaln*, 2024.

[3] Eric Wengrowski and Kristin Dana. Light field messaging with deep photographic steganography. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:1515–1524, 2019.

[4] Fabrizio Guillaro, Davide Cozzolino, Avneesh Sud, Nicholas Dufour, and Luisa Verdoliva. TruFor: Leveraging All-Round Clues for Trustworthy Image Forgery Detection and Localization. pages 20606–20615, 2023.

[5] Naima Otberdout, Mohamed Daoudi, Anis Kacem, Lahoucine Ballihi, and Stefano Berretti. Dynamic facial expression generation on hilbert hypersphere with conditional wasserstein generative adversarial nets. *arXiv*, pages 1–21, 2019.

[6] Zhen Zhu, Tengteng Huang, Baoguang Shi, Miao Yu, Bofei Wang, and Xiang Bai. Progressive pose attention transfer for person image generation. *arXiv*, 2019.

[7] Wei Pan, Anna Zhu, Xinyu Zhou, Brian Kenji Iwana, and Shilin Li. Few shot font generation via transferring similarity guided global style and quantization local style. pages 19506–19516, 2023.

[8] Tanmoy Mondal, Ricardo Mendoza, and Lucas Drumetz. Physics informed and data driven simulation of underwater images via residual learning, 2024.

[9] Jiangxin Dong and Jinshan Pan. Physics-Based Feature Dehazing Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12375 LNCS:188–204, 2020.

[10] Michael Ramamonjisoa, Michael Firman, Jamie Watson, Vincent Lepetit, and Daniyar Turmukhambetov. Single Image Depth Prediction with Wavelet Decomposition. pages 11084–11093, 2021.

[11] Tanmoy Mondal, Reza Akbarinia, and Florent Masseglia. knn matrix profile for knowledge discovery from time series. *Data Mining and Knowledge Discovery*, 37:1055–1089, 5 2023.

[1] `https://mondal-tanmoy.github.io/files/pdf/conf/KDD_Ver_1.pdf` `https://mondal-tanmoy.github.io/files/pdf/journal/kNN_Matrix_Profile.pdf`