Robust Day-Night Image Matching Across Extreme Illumination Variations: A Comparative Study of Deep Learning and Classical Methods

Biyun Yan

Department of Electrical and Computer Engineering, Worcester,
Polytechnic Institute, 100 Institute Rd, Worcester, 01609, MA, USA

Abstract: ***The clinical applicability of Magnetic Resonance Imaging (MRI) is often constrained by prolonged scanning durations, particularly in emergency scenarios. To address this, this study conducts a comparative analysis between a traditional greedy algorithm, Orthogonal Matching Pursuit (OMP), and a deep learning alternative known as Generative Adversarial Networks for compressed sensing (GANCS). Empirical evidence suggests that GANCS outperforms the classical method by delivering higher fidelity images from highly undersampled data, while also ensuring rapid processing speeds following the training phase. obust image matching across extreme illumination conditions remains a critical challenge in computer vision, paricularly for sustainability-related applications such as smart city monitoring, urban safety, and environmental surveillance. This study evaluates the performance of local feature matching techniques under severe day–night illumination variations using images from the AMOS dataset. I compare a deep learning–based approach, MatchNet, with a classical template-matching method based on distance transforms and normalized cross-correlation. Experimental results demonstrate that while deep learning models exhibit strong feature repeatability, they struggle to achieve reliable matching precision under extreme illumination changes. In contrast, the classical approach shows superior mean average precision, particularly for cross-domain image retrieval tasks. The findings highlight the importance of task-specific modeling choices and suggest directions for improving illumination robust feature matching in sustainable vision systems***

Keywords: Compressed sensing; MRI reconstruction; Orthogonal Matching Pursuit; Generative Adversarial Networks; Deep learning; Undersampled MRI; Image reconstruction.

1. INTRODUCTION

Mision-based monitoring systems play an increasingly important role in sustainable urban infrastructure, enabling applications such as traffic analysis, environmental observation, and public safety monitoring. A persistent challenge in these systems is robust image matching across drastic illumination changes, particularly between daytime and nighttime conditions. Illumination variations in urban scenes are often spatially non-uniform due to artificial lighting, shadows, and reflective surfaces. As observed in the AMOS dataset, feature detector repeatability for day–night image pairs is significantly lower than for day–day or night–night pairs. Moreover, high repeatability does not necessarily translate to strong matching performance. Prior work has shown that learned detectors optimized for illumination invariance may still suffer from low matching precision and recall, indicating that repeatability alone provides only a loose upper bound on achievable correspondence quality. This study investigates both deep learning–based and classical image matching methods to better understand their respective strengths and limitations under extreme illumination changes [1].

2. METHOD

Implement the MatchNet framework proposed in [2] as the basis for our neural network architecture. The main contributions of the original work can be summarized as follows:

2.1 Concept

Orhe MatchNet architecture is composed of two main components: a feature network and a metric network. The feature network is inspired by AlexNet [3], which has demonstrated strong performance in object recognition tasks. Its primary function is to extract discriminative feature representations from image patches. Rectified Linear Unit (ReLU) activations are employed in all convolutional layers to introduce non-linearity. The metric network is designed to model the similarity between pairs of feature representations using three fully connected layers with ReLU activation. In this work, the original FC3 and softmax layers are replaced with a sigmoid activation. The input to the metric network is the concatenation of two feature vectors, and the output represents the estimated probability that the two corresponding image patches match or do not match. A two-tower architecture with shared parameters, analogous to a Siamese network, is adopted during model construction. During training, two identical feature networks feed [3]

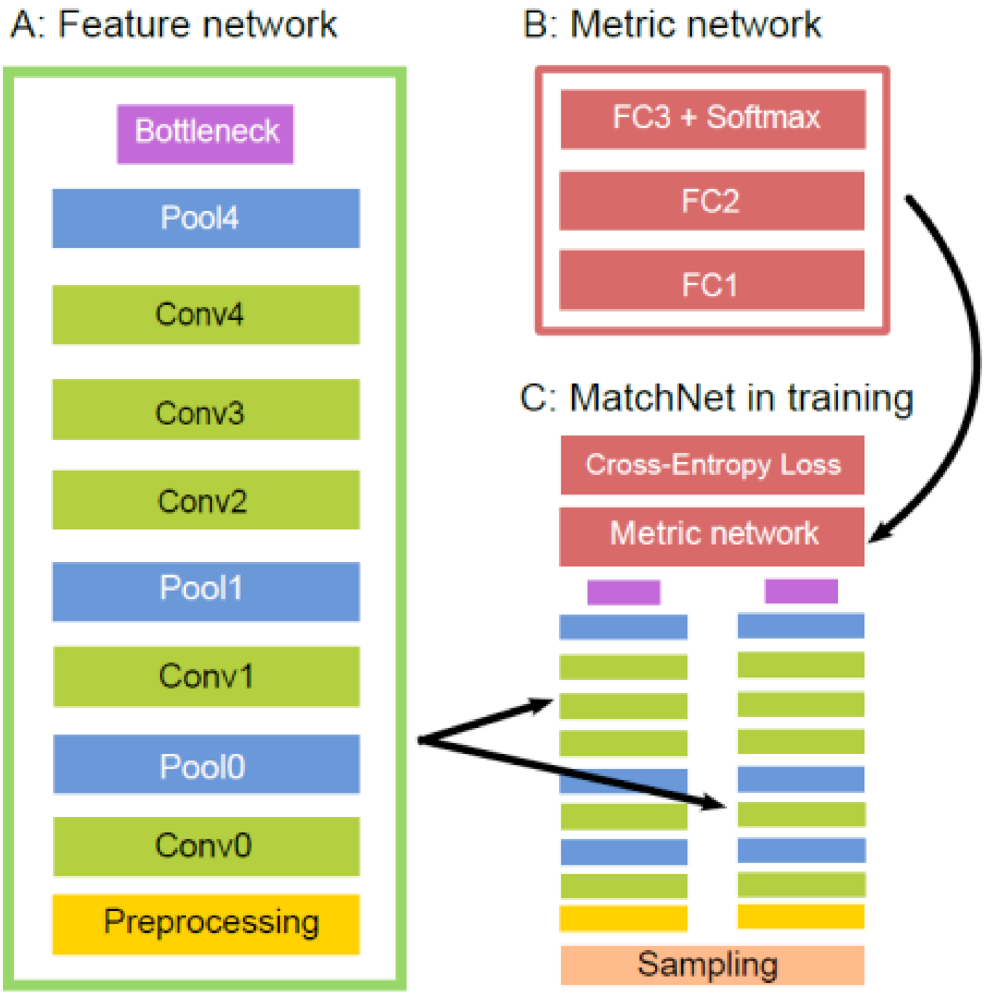


Figure 1: MatchNet

Shared comparison network, and gradient updates from both branches are applied to the same set of parameters. Compared with stereo reconstruction networks, MatchNet incorporates additional convolutional layers and max-pooling operations to improve robustness to scale variations. To reduce the dimensionality of the feature representation and mitigate overfitting, a bottleneck fully connected layer [3] is introduced between the max-pooling output (of size $25 \times 25 \times 24$) and the final feature embedding. Prior to training, all RGB images are converted to grayscale and normalized using dataset-specific statistics, where pixel intensity values $x \in [0, 255]$ are transformed to $(x - 127.723)/73.959$ [4-11]

2.2 Data Processing

Limit the appearance of such artifacts and noise, a modifi-cation of AHE called Contrast Limited AHE can be used. The amount of contrast enhancement for some intensity is directly proportional to the slope of the CDF function at that intensity level. Hence contrast enhancement can be limited by limiting the slope of the CDF. The slope of CDF at a bin location is determined by the height of the histogram for that bin. Therefore if I limit the height of the histogram to a certain level I can limit the slope of the CDF and hence the amount of contrast enhancement [13-19].

Figure 2: Image after CLAHE processing and original image

2.3 Metrics of Result

Average precision is a popular metric in measuring the accuracy of object detectors. AP computes the average precision value for recall value over 0 to 1. Precision measure how accurate is our predictions. Recall measures how good I find all the positives. Their mathematical definitions are Q stands for query images.

$$Precision = \frac{Truepositive}{Trueposition+Falsepositive}$$

$$Recall = \frac{Truepositive}{Trueposition+Falsenegative}$$

$$MAP = \frac{\sum_{q=1}^{Q} AueP(q)}{Q}$$

First build our data set, dividing training data set 15 scenarios into one positive data set and negative data set. Positive data set is labeled by one, while the negative one is zero. The input data is a random pair of image patch and output is one or zero labels. Second, train our model in epoch 500 times and save the best model (normally have a best accuracy) for further testing. Our loss function is the contrast loss function, which use Euclidean distance to measure the similarity between two pictures. When the distance tends to increase, that means two patches of images are highly from different classes. On the other hand, decreasing distance implies the similarity of two patches. See Figure 3 the process of training our model

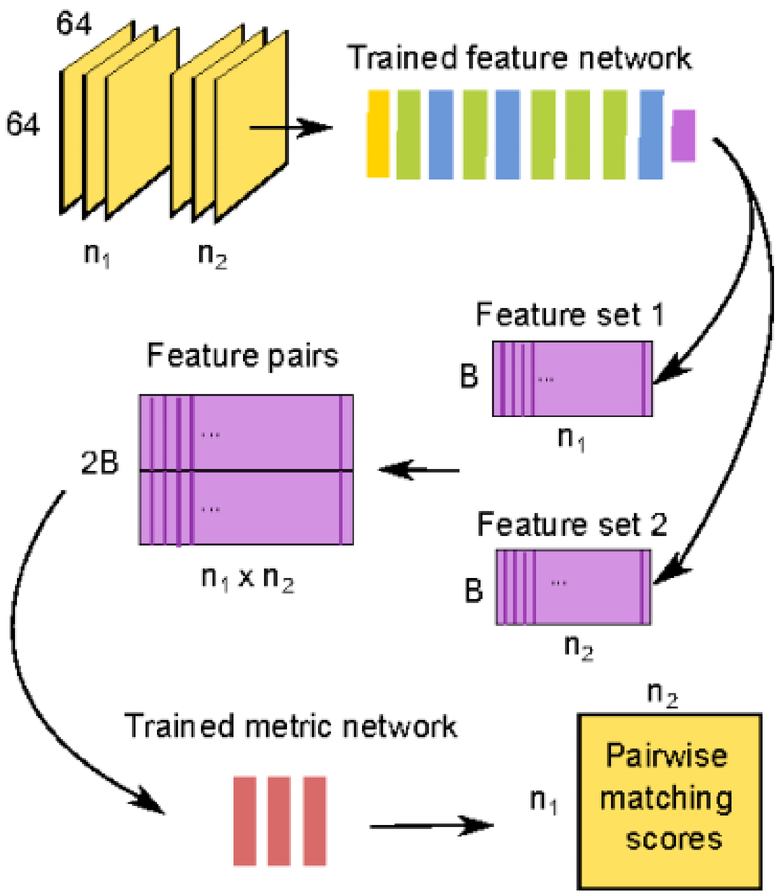
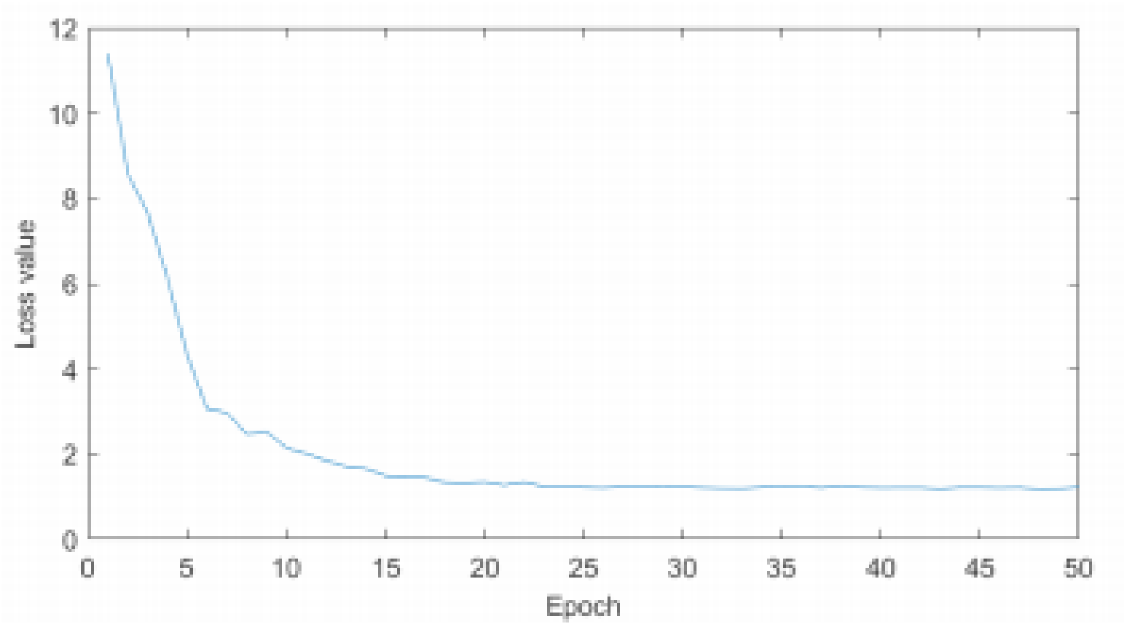
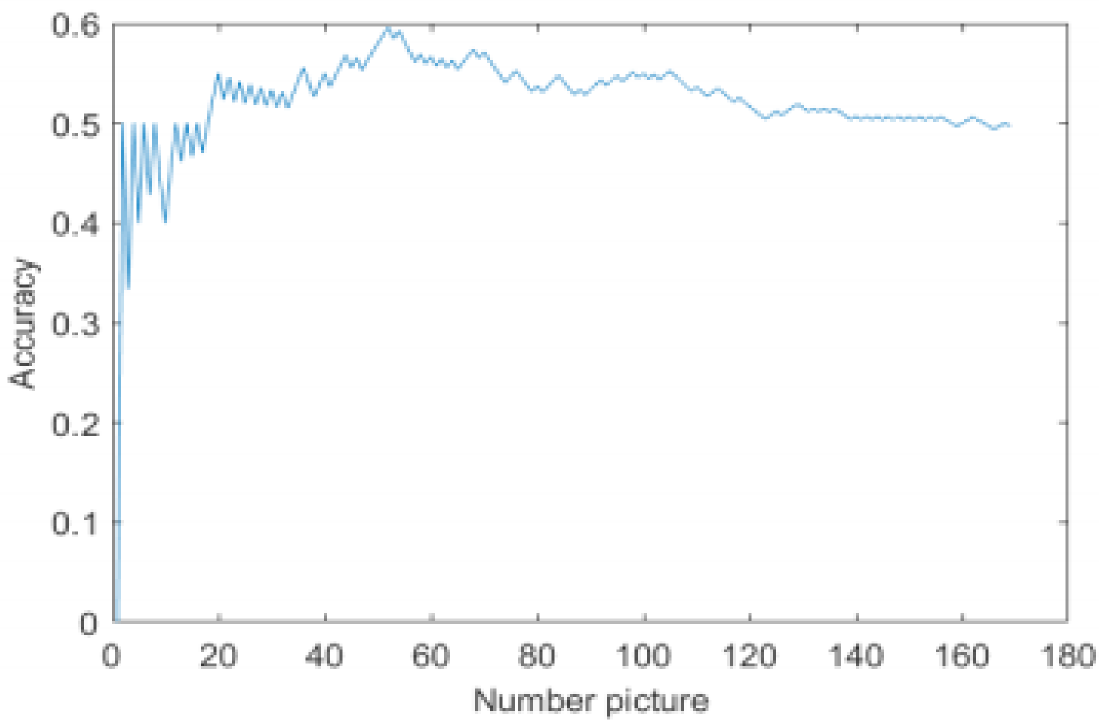


Figure 3: MatchNet is disassembled during prediction.

At the time, I have our best results, testing by testing data set and calculating the metrics of our model





3. DISTANCE TRANSFORM WITH TEMPLATE MATCHING

The following section will describe the methodology of the classical method of trying to solve this problem. The general steps for each section of the process can be seen in Figure 6.

Since, the dataset contained very low light and RGB images, the first step was to process the images. Firstly, the algorithm converted the RGB to grayscale by removing the saturation, but keeping the luminance. Secondly, Histogram Equalization was performed on both input images to increase the contrast in the images. Thirdly, denoising was performed through self guided filtering. Lastly, image sharpening using the unsharp masking technique was performed to enhance the feature points, most notably the edges in the image. The resulting output from an example image can be seen in Figure 7. The two processed images are then transformed to edge, binary value, images using the Canny Edge detection algorithm. The standard deviation of the filter was set to 2 and the low and high thresholds were set to .000001 and .4 respectively. The values were found through testing with training datasets to acquire robust features and minimize noise capturing. Distance transformation on the normalized edge images were computed using the Euclidean distance method. The images are normalized to a range of 10 for better comparison. Figure 8 presents the resulting outputs of this section.

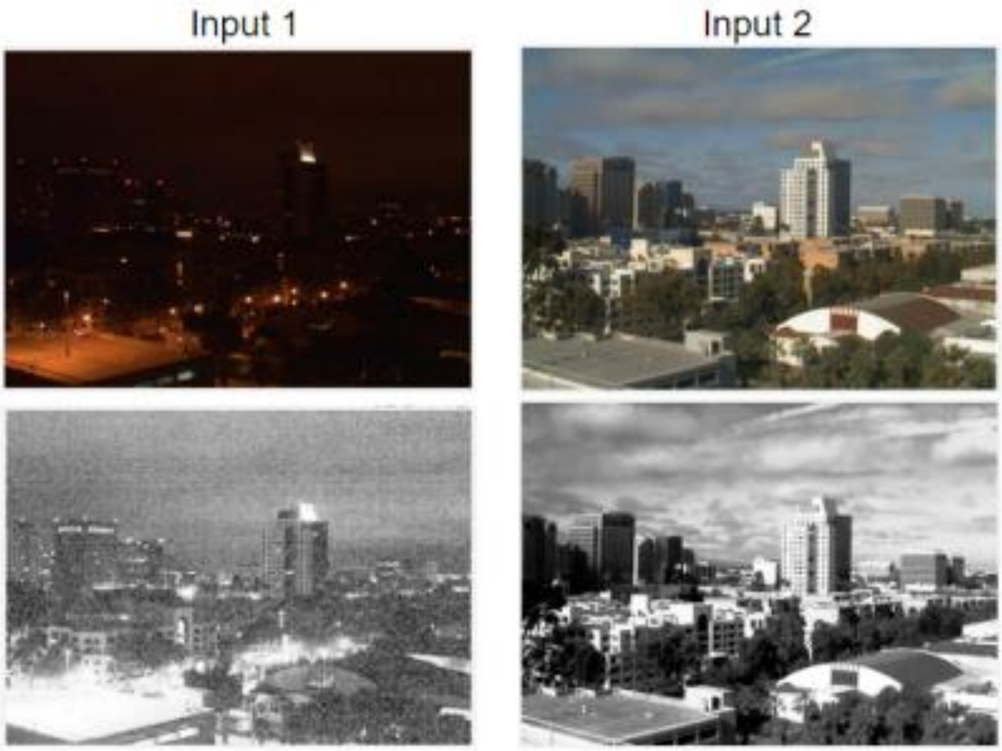


Figure 7: Example image enhancement results

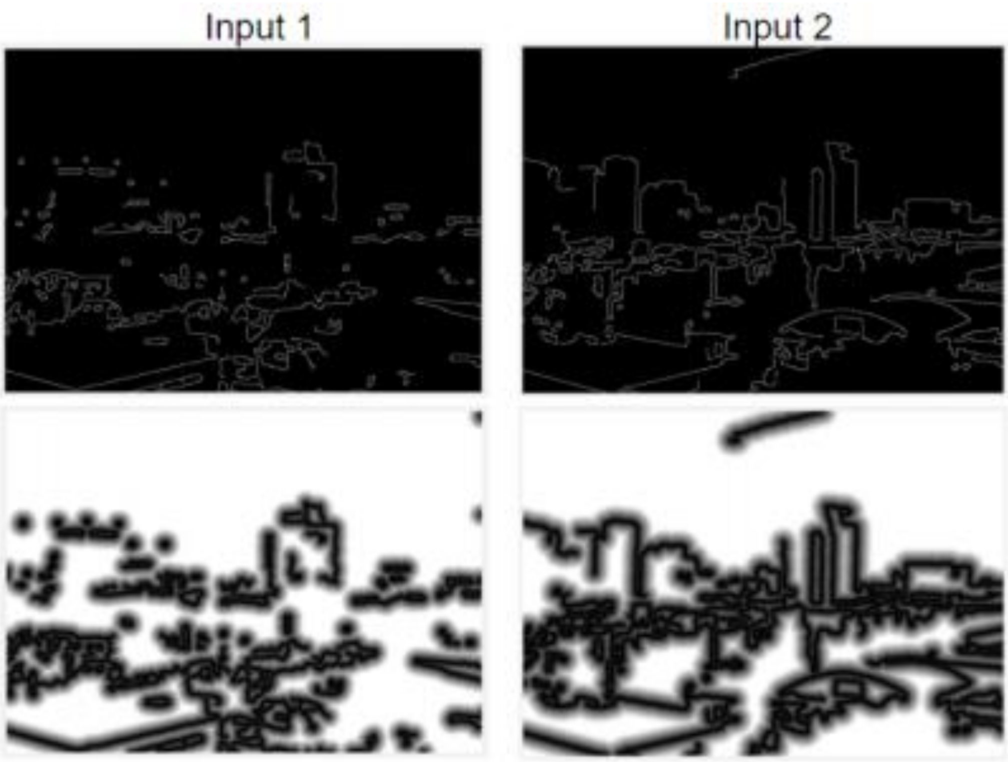


Figure 8: Example image edge detection and distance transform results

One processed input from the previous step will be split into quadrants. Each quadrant will be represented by a value that is the average of the entire patch. The quadrant with the low est value, representing the quadrant with the most edge features, will be chosen as the template for the next phase of template matching. Figure 9 displays an example of a resulting patch determined to be the template. Template matching was performed using normalized cross correlation.The output of the method determined the correlation coefficient mapping. The final value determining the score of the two input images was the maximum correlation coefficient found in the entire mapping. Figure 7

4. CONCLUSION

In the classical method, results indicated that mean average precision (mAP) was highest when no image enhancement was applied — that is, without histogram equalization, image sharpening, or denoising. This outcome is likely attributable to the fact that histogram equalization introduced considerable noise into the images, causing the edge map to capture spurious artifacts alongside genuine features. While image enhancement did improve the perceptual quality of nighttime images, daytime images subjected to the same enhancement pipeline generated additional noise unnecessarily, as they do not require histogram adjustment in the first place. In future work, more sophisticated denoising techniques or adaptive contrast measurement methods could be employed to determine whether enhancement is warranted on a per-image basis, rather than applying it uniformly.

For the deep learning method, the relatively low accuracy may be partly attributed to the use of contrast images as the sole input modality, which limits the richness of information available to the model. In the case of the MatchNet model, performance fell short of the classical methods, likely because converting input images into edge representations discards potentially discriminative information that would otherwise aid in matching.

To further assess the robustness of both approaches, a more diverse set of image scenes should be incorporated into the testing dataset. In the classical method, the algorithm performed well when matching nighttime images to nighttime images and daytime images to daytime images, but struggled considerably with cross-condition matching between day and night captures of the same scene. This limitation is clearly reflected in the substantial drop in precision when comparing top-10 returns against all returns.

In this study, I evaluated the performance of local feature-based techniques for day-night image matching. For template matching, cross-correlation was computed across day-night image pairs to quantify their similarity, where higher scores indicate greater similarity. The classical method achieved higher mAP scores than the deep learning approach, which is consistent with the fact that cross-correlation is well suited to image retrieval tasks. The deep learning method, formulated as a binary classification problem, yielded lower-than-expected mAP scores. Nevertheless, there remains considerable potential for improving the deep learning model through several directions: augmenting the training dataset via mirroring or rotation, adopting a triplet loss function to better capture relative similarity, or redesigning the model as a dedicated detector architecture to achieve greater robustness under severe illumination changes.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp.529-551, April 1955.

[2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] Qian, C., Guo, Y., Mo, Y., & Li, W. (2025). WeatherDG: LLM-Assisted Procedural Weather Generation for Domain-Generalized Semantic Segmentation. IEEE Robotics and Automation Letters, 10(6), 5919 – 5926. https://doi.org/10.1109/lra.2025.3559821

[4] Qian, C., Guo, Y., Li, W., & Markkula, G. (2025). Weathergs: 3D Scene Reconstruction in Adverse Weather Conditions Via Gaussian Splatting. In 2025 IEEE International Conference on Robotics and Automation (ICRA) (pp. 185–191). IEEE. 2025 IEEE International Conference on Robotics and Automation (ICRA). https://doi.org/10.1109/icra55743.2025.11128699

[5] Wu, W., Guo, Y., Li, Qi, & Jia, C. (2024). Exploring the potential of large language models in identifying metabolic dysfunction-associated steatotic liver disease: A comparative study of non-invasive tests and artificial intelligence-generated responses. Liver International, 45(4). https://doi.org/10.1111/liv.16112

[6] Qian, C., Li, W., Guo, Y., & Markkula, G. (2025). WeatherEdit: Controllable Weather Editing with 4D Gaussian Field (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2505.20471
[7] Guo, Y., Qian, C., Mo, Y., & Sangpetch, A. (2025). GaussianSlicer: Efficient Surface Reconstruction from Cross-sectional Slices with Gaussian Splatting. In ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1–5). IEEE. ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/icassp49660.2025.10890834
[8] J. Li, T. B. Culver, P. P. Persaud, and J. M. Hathaway, "Developing nitrogen removal models for stormwater bioretention systems," Water Research, vol. 243, p. 120381, 2023.
[9] J. Li and T. B. Culver, "Review of process-based nitrogen model for agricultural fields with implications for nitrogen simulations in stormwater BMPs," Environmental Modelling & Software, vol. 151, p. 105363, 2022.
[10] Li, T. B. Culver, C. R. Burgis, W. Zhang, and J. A. Smith, "Validating Nitrogen Removal Models with Field Bioretention Data," Journal of Environmental Engineering, vol. 150, no. 8, p. 04024037, 2024.
[11] Li, "Nitrogen Removal Models for Stormwater Bioretention Systems," Ph.D. dissertation, University of Virginia, 2023.
[12] Liang, J., Wang, Z., Ma, Z., Li, J., Zhang, Z., Wu, X., & Wang, B. (2024). Online training of large language models: Learn while chatting. arXiv preprint arXiv:2403.04790.
[13] Wang, Z., Su, J., Zhou, M., Zeng, H., Jia, M., Lv, X., ... & Zhang, D. (2025). SheetBrain: A Neuro-Symbolic Agent for Accurate Reasoning over Complex and Large Spreadsheets. arXiv preprint arXiv:2510.19247.
[14] Ge, W., Wang, Z., Wang, P., Liang, J., Cai, Z. G., Mai, Z., & Wang, B. (2024). Towards Gamifying Interactive Language Learning using Large Language Models for Children.
[15] Wang, Z., Zhang, Q., Liu, T., & Li, C. (2024). Analyzing Financial News Sentiment with NLP to Forecast Market Trends. International Journal of Engineering and Management Research, 14(5), 6-11.
[16] Rao, Jiarui, et al. "Optimizing Stock Market Return Forecasts with Uncertainty Sentiment: Leveraging LLM-based Insights." Proceedings of the 2024 5th International Conference on Big Data Economy and Information Management. 2024.
[17] Rao, Jiarui, and Jionghao Lin. "Ramo: Retrieval-augmented generation for enhancing moocs recommendations." arXiv preprint arXiv:2407.04925 (2024).
[18] Rao, Jiarui, Qian Zhang, and Xinqiu Liu. "Applications Analyzing E-commerce Reviews with Large Language Models (LLMs): A Methodological Exploration and Application Insight." Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 7.01 (2024): 207-212.
[19] Zhang, Qian, et al. "Sea MNF vs. LDA: Unveiling the power of short text mining in financial markets." International Journal of Engineering and Management Research 14.5 (2024): 76-82.
[20] Rao, Jiarui, et al. "Integrating Textual Analytics with Time Series Forecasting Models: Enhancing Predictive Accuracy in Global Energy and Commodity Markets." Innovations in Applied Engineering and Technology (2023): 1-7.
[21] Zhang, Qian, and Jiarui Rao. "Enhancing Financial Forecasting Models with Textual Analysis: A Comparative Study of Decomposition Techniques and Sentiment-Driven Predictions." Innovations in Applied Engineering and Technology (2022): 1-6.
[22] Lin, Jionghao, et al. "Automatic large language models creation of interactive learning lessons." European Conference on Technology Enhanced Learning. Cham: Springer Nature Switzerland, 2025.
[23] Peng, Jingyang, et al. "Automated bias assessment in ai-generated educational content using ceat framework." International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland, 2025.
[24] Rao, Jiarui, and Qian Zhang. "Deconstructing Digital Discourse: A Deep Dive into Distinguishing LLM-Powered Chatbots from Human Language." Journal of Theory and Practice in Education and Innovation 2.2 (2025): 18-25.
[25] Deng, Xiaoxiao. "Enhancing Neural Network Performance on Tabular Data via Knowledge Distillation and RankGauss Transformation." 2025 6th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE, 2025.
[26] Deng, Xiaoxiao. "Graph Inference Towards ICD Coding." 2025 3rd International Conference on Artificial Intelligence and Automation Control (AIAC). IEEE, 2025.
[27] Zi, Yun, and Xiaoxiao Deng. "Joint modeling of medical images and clinical text for early diabetes risk detection." Journal of Computer Technology and Software 4.7 (2025).
[28] Xu, Ting, et al. "Clinical NLP with attention-based deep learning for multi-disease prediction." 2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC). IEEE, 2025.