An Intelligent Multi-Model Framework for Online Recruitment Fraud Detection Using Trust Score and Explainable AI: A Review

Dr. S. S. Khatal[1], Dr. P. S. Gholap[2], Miss. Tattu Minal Goraksh[3], Miss. Nawale Mahima Nanasaheb[4], Mr. Maval Vivek Sanjay[5]

[1,2]*Assistant Professor, Department of Computer Engineering Sharadchandra Pawar College of Engineering Dumberwadi, Pune (MH) India*

[3,4,5]*Student, Department of Computer Engineering Sharadchandra Pawar College of Engineering Dumberwadi, Pune (MH) India*

***Abstract*—Online recruitment fraud has emerged as a significant cybersecurity concern, deceiving job seekers through fake advertisements, phishing emails, and fraudulent offers on digital platforms. This paper presents an Advanced Fake Job Post Detection System that leverages machine learning and trust-based analysis to accurately identify deceptive job postings in real time. The proposed framework integrates multiple models Multilayer Perceptron (MLP), Passive Aggressive Classifier, Gradient Boosting, and K-Nearest Neighbors allowing flexible and comparative prediction for enhanced reliability. Unlike traditional systems limited to static datasets, this model processes live job data and email inputs, including attachments, to determine authenticity. The system performs comprehensive data preprocessing involving text cleaning, encoding, and feature selection to focus on key attributes such as job description, company logo, education, and experience requirements. A Trust Score mechanism is introduced to quantify the credibility of each post, classifying them as suspicious, uncertain, or verified real, while an explanation report provides reasoning for each prediction. Furthermore, trend analysis visualizes the distribution of fake posts across platforms and job categories. The experimental results demonstrate that the proposed model achieves improved accuracy, faster processing, and practical applicability, offering a robust, transparent, and user-centric solution for mitigating online recruitment frauds and safeguarding job seekers.transparent, and user-centric solution for mitigating online recruitment frauds and safeguarding job seekers.**

***Index Terms*—Online Recruitment Fraud, Fake Job Post Detection, Machine Learning, Trust Score, Real-Time Classification, Feature Selection, Text Analytics, Gradient Boosting, Explainable AI, Cybersecurity.**

I. INTRODUCTION

Online recruitment has become one of the most widely used channels for connecting employers and job seekers in the digital era. With the rapid expansion of job portals, social media-based hiring, and email-based recruitment, millions of employment opportunities are shared online every day. However, this growth has been accompanied by a surge in fake job postings and recruitment scams, posing serious financial and psychological risks to applicants [1], [2]. Fraudsters exploit popular job platforms by posting deceptive advertisements to collect personal data, extract money, or engage in identity theft [3]. Traditional monitoring mechanisms and manual verification methods are insufficient to detect such fraudulent content, making automated fake job post detection a crucial need in the modern employment ecosystem [4].

Existing research on online recruitment fraud detection has mainly focused on basic machine learning models like Logistic Regression, Naive Bayes, or Support Vector Machines [5], [6]. Although these models offer decent performance, they often rely on static datasets and are unable to handle real-time data or adapt to new patterns of deception [7]. Moreover, many conventional systems fail to consider feature selection and data preprocessing, leading to reduced model accuracy and slower

prediction times [8]. Another limitation is that they rarely incorporate explainability or trust-based evaluation, leaving users uncertain about why a specific job post was flagged as fake or genuine [9].

The proposed system addresses these challenges by introducing a multi-model intelligent framework that integrates four different algorithms—Multilayer Perceptron (MLP), Passive Aggressive Classifier, Gradient Boosting, and K-Nearest Neighbors (KNN)—for comparative and ensemble prediction [10]. This diversity allows the system to capture nonlinear text patterns, structured attributes, and hidden relationships between job post features, enhancing detection accuracy and robustness. Furthermore, a Trust Score mechanism is implemented to quantify the reliability of each prediction, assigning percentage-based confidence levels to classify posts as very suspicious, potentially suspicious, or verified real [11], [12]. By incorporating this metric, the model not only provides a binary decision but also a human interpretable confidence level, improving user trust and transparency.

Data preprocessing plays a vital role in ensuring the stability and quality of predictions. The proposed framework performs comprehensive preprocessing steps such as missing value handling, tokenization, normalization, and encoding to eliminate inconsistencies from the input dataset [13]. It emphasizes critical features like job description, experience requirement, educational qualification, company logo, and email domain—attributes found to be highly influential in distinguishing real from fake postings [14]. Additionally, natural language processing (NLP) techniques are employed to analyze linguistic cues, sentiment, and urgency indicators in job descriptions, as fraudulent recruiters often use persuasive or emotional language to mislead applicants [15].

Another key innovation of this work is the Explainable AI (XAI) component, which generates detailed reports justifying each prediction. The system highlights top contributing factors—such as suspicious keywords, missing corporate information, or unverifiable contact details—providing users with a clear rationale behind every classification [16]. This interpretability is critical in promoting responsible AI use, particularly in high-stakes scenarios where incorrect classifications could lead to missed employment opportunities or exposure to fraud [17]. Moreover, the system visualizes trends and statistical insights, identifying patterns like which platforms, industries, or regions exhibit higher rates of recruitment fraud [18].

The proposed system is not confined to academic experimentation but designed for real world deployment, accessible via a web dashboard or browser extension. It processes both online job posts and recruitment emails, including attachments, to assess authenticity in real time [19]. Experimental results demonstrate superior performance in terms of accuracy, precision, and response time compared to traditional single-model approaches. By integrating multi-model learning, trust scoring, and explainable reporting, the system offers a scalable, 2 transparent, and user-friendly solution to safeguard job seekers from fraudulent activities in the online employment market[20].

II. MOTIVATION

The growing number of fake job postings on online platforms poses a serious threat to job seekers, leading to financial losses, identity theft, and emotional distress. Despite advancements in automation and digital recruitment, many detection systems still rely on limited datasets and static models, making them ineffective against evolving fraud tactics. Hence, there is a strong need for an intelligent, real-time, and explainable system that not only identifies fraudulent job posts with high accuracy but also provides users with trust scores and reasons for classification. This motivation drives the development of an advanced, user-centric fraud detection framework that ensures safer online recruitment experiences.

Goals and Objectives

1. To study various machine learning and deep learning algorithms for effective detection of fake job postings in online recruitment systems.
2. To study the importance of key features such as job description, logo, experience, and education in improving classification accuracy.
3. To study real-time data processing methods for analyzing live job posts and identifying fraudulent patterns dynamically.

4. To study the implementation of trust score mechanisms that help users evaluate the authenticity of job posts with transparency.
5. To study the integration of explainable AI techniques that provide clear reasons behind each prediction, enhancing user trust and system reliability.

Scope

The scope of this project focuses on designing and implementing an intelligent Online Recruitment Fraud Detection System capable of identifying fake job postings across multiple online platforms in real time. The system utilizes advanced machine learning algorithms such as MLP, Passive Aggressive, Gradient Boosting, and K-Neighbors to enhance detection accuracy and adaptability. It integrates email verification, trust scoring, and explanation based reporting to make predictions transparent and user-friendly. The system also includes a data preprocessing pipeline for cleaning, encoding, and normalizing input data, ensuring reliable outcomes. Beyond academic research, this solution is designed for practical deployment, enabling job seekers, students, and recruitment agencies to verify the authenticity of job posts and prevent cyber scams. Future scalability includes integration with APIs of job portals, real-time dashboards, and multilingual support to strengthen its global applicability.

III. EXISTING SYSTEM

The increasing digitization of recruitment processes has made online job portals and professional networking platforms vulnerable to fraudulent activities. Researchers have explored various machine learning and deep learning methods to detect fake job postings, focusing on automation, feature extraction, contextual analysis, and real-time detection. This section discusses the key contributions and limitations of previous works in this domain.

M.S.M. Anitha, Y. Naga Malleswarao, and Puppala Ajay [1] introduced a deep learning based detection system using Convolutional Neural Networks (CNNs) and Multi-Layer Perceptrons (MLPs). Their study achieved 95.1% accuracy and provided a web-based interface for real-time detection. Although the model demonstrates strong performance, it lacks a thorough feature engineering process and comparative analysis with other models. Moreover, the application of CNNs to text-based data, while innovative, was not fully optimized for linguistic representations. Despite these limitations, this work establishes a foundation for practical and deployable recruitment fraud detection systems.

In a pioneering effort, Bandar Alghamdi and Fahad Alharby [2] developed an intelligent ensemble model employing Random Forest for classification and Support Vector Machine (SVM) for feature selection, trained on the Employment Scam Aegean Dataset (EMSCAD). Achieving an impressive 97.41% accuracy, their study emphasized the significance of feature importance, with attributes such as company profile, logo, and industry serving as key indicators of fraud. This work set a strong baseline for subsequent research by proving that ensemble learning significantly enhances detection performance in online recruitment fraud (ORF) contexts.

Syed Mahbub, Eric Pardede, and A.S.M. Kayes [3] extended the research frontier by examining contextual and regional features in recruitment fraud detection within the Australian job industry. Using a localized dataset derived from Gumtree, they integrated contextual signals such as company registry validation, web domain existence, and Named Entity Recognition (NER)-based entity matching. The inclusion of these features improved the model's predictive accuracy to 91.8%, showcasing the importance of domain-specific and localized data. This study provided crucial insights into the role of contextual data in fraud detection models, emphasizing adaptability across geographic regions. Dr. P. Vara Prasad, N. Lakshmi Sravya, M. Sree Kavya, V. Kavitha, and Shaik [4] proposed a deep learning-based hybrid model combining Long Short-Term Memory (LSTM), CNN, and Bidirectional Encoder Representations from Transformers (BERT). Their framework integrates real-time web scraping and anomaly detection for identifying fraudulent job listings. The system achieved 92% accuracy and demonstrated robustness across various job categories. While the study highlights the potential of deep learning for capturing complex linguistic patterns, it lacks a detailed discussion on class imbalance handling and computational optimization. Nevertheless, its emphasis on real-time adaptability marks a

significant step toward operational scalability. In a comprehensive comparison of traditional machine learning classifiers, Shawni Dutta and Prof. Samir Kumar Bandyopadhyay [5] examined both single classifiers (Naive Bayes, MLP, KNN, Decision Tree) and ensemble approaches (Random Forest, AdaBoost, Gradient Boosting). Their results, obtained using the EMSCAD dataset, revealed that Random Forest achieved the highest accuracy of 98.27%. This study confirmed that ensemble techniques outperform single classifiers in terms of precision, recall, and generalization. Furthermore, their methodological rigor and reliance on standard datasets make the results replicable and trustworthy, serving as a benchmark for future research. A major advancement in this field is represented by Fraud-BERT, proposed by Khushboo Taneja, Jyoti Vashishtha, and Saroj Ratnoo [6]. This transformer-based model leverages 4 BERT's contextual embeddings to overcome the limitations of bag-of-words and TF-IDF approaches. Fraud-BERT achieved 99% accuracy and a 0.93 F1-score on the imbalanced EMSCAD dataset without the need for resampling, demonstrating its strong contextual understanding. Moreover, the study discusses the ethical implications and computational challenges of deploying transformer-based fraud detection models. It establishes a new benchmark in accuracy and context sensitivity, highlighting the growing trend toward deep contextual and explainable AI systems. A comparative review of these studies indicates that although deep learning and ensemble methods have achieved remarkable accuracy, most systems are dataset-dependent and lack cross-domain adaptability. Few models offer interpretability or transparency, which are essential for real-world use. Furthermore, real-time detection, multi-model flexibility, and trust-based scoring remain underexplored. The proposed work in this paper addresses these gaps by introducing a multi-model intelligent framework that combines machine learning adaptability, explainable AI, and a trust-based scoring mechanism to enhance detection reliability and user trust.

IV. PROPOSED SYSTEM

The proposed Online Recruitment Fraud Detection System is an advanced, intelligent, and real-time platform designed to identify fake job postings using multiple machine learning algorithms and explainable decision-making mechanisms. The system is structured to improve accuracy, interpretability, and user safety by analyzing essential features from job postings and emails while providing actionable insights through trust scores and detailed reports. The system architecture comprises several functional modules — data acquisition, preprocessing, feature selection, model training and testing, trust score calculation, result explanation, and visualization. Each module contributes to ensuring that the model operates efficiently in real-world environments, providing real-time fraud analysis with transparency and reliability.

1. Data Acquisition:

The system collects data from verified recruitment platforms, email attachments, and existing datasets such as the Employment Scam Aegean Dataset (EMSCAD). It also supports real-time web scraping to fetch new job postings. This approach ensures that the model continuously learns from diverse and updated data sources, improving adaptability to emerging fraud patterns.

2. Data Preprocessing:

Raw job data often contain missing values, duplicate records, and irrelevant features. The preprocessing module handles these issues by applying data cleaning, label encoding, normalization, and text preprocessing (tokenization, stop-word removal, and lemmatization). By standardizing and transforming the data, the system ensures high-quality input for model training, reducing noise and improving accuracy.

3. Feature Selection:

The system focuses only on critical features such as job title, job description, company logo, education, experience, employment type, and salary range. Feature importance is determined using statistical methods and model-based techniques, ensuring only relevant inputs are used for prediction. This selective approach improves model speed and reduces overfitting, resulting in more stable outcomes.

4. Model Training and Prediction:

The core of the proposed system lies in the integration of four powerful machine learning algorithms — Multi-Layer Perceptron (MLP), Passive Aggressive Classifier, Gradient Boosting, and K-Nearest Neighbors (KNN).

- MLP provides deep pattern recognition for complex feature relationships.
- Passive Aggressive handles online learning effectively, adapting quickly to new data.
- Gradient Boosting offers high accuracy through iterative error minimization.
- KNN ensures simple, instance-based classification for explainable results. Users can select their preferred model, allowing flexibility and transparency in predictions.

5. Trust Score Mechanism:

Once a prediction is made, the system generates a Trust Score that quantifies the authenticity of a job post:

- 20% – Very Suspicious (likely fake)
- 60% – Doubtful/Needs Verification
- 95% – Verified Real Post

The trust score is computed based on probabilistic confidence levels, feature weight contributions, and cross-model validation results. This feature empowers users to make informed decisions before applying for jobs.

6. Reasoning and Report Generation:

To enhance transparency, the system provides a detailed reasoning report explaining why a job post was classified as real or fake. The report includes insights such as missing company logos, suspicious job descriptions, unrealistic salary offers, or mismatched contact details. These explanations make the system explainable and user-trustworthy, addressing one of the major challenges in AI-based fraud detection.

7. Visualization and Trend Analysis:

The system features a visual analytics dashboard displaying trends and insights, such as which job categories or locations have the highest fraud rates. Graphs like ―Fake vs Real Posts Over Time,‖ ―Category-wise Fraud Distribution,‖ and ―Model Accuracy Comparison‖ help in understanding the evolving fraud landscape. This analytical view aids recruiters and researchers in identifying high-risk zones in the online recruitment domain.

8. Email Verification Module:

In addition to job posts, the system analyzes emails with attachments to determine their legitimacy. It checks metadata, email headers, sender authenticity, and attached document structure to detect phishing or fraudulent communications, thus offering a complete recruitment safety ecosystem.

9. Deployment and User Interaction:

The proposed system can be deployed as a web-based application where users upload job details or email content for verification. The model processes input data in real time and returns results instantly. The architecture supports scalability, meaning more datasets, models, and features can be integrated in the future to handle global job fraud detection needs.

In summary, the proposed system provides a multi-model, explainable, and real-time solution to online recruitment fraud detection. By combining robust preprocessing, intelligent feature engineering, flexible model selection, and transparent reporting, the system ensures high performance and reliability. It not only benefits job seekers by safeguarding their applications but also assists recruiters and employment platforms in maintaining credibility and trust in digital recruitment ecosystems.

V. SYSTEM DESIGN

The system architecture is designed with multiple layers, each responsible for a specific function — from data collection to visualization. The workflow begins with collecting job post or email data, followed by preprocessing, feature extraction, model prediction, and result visualization with a trust score.

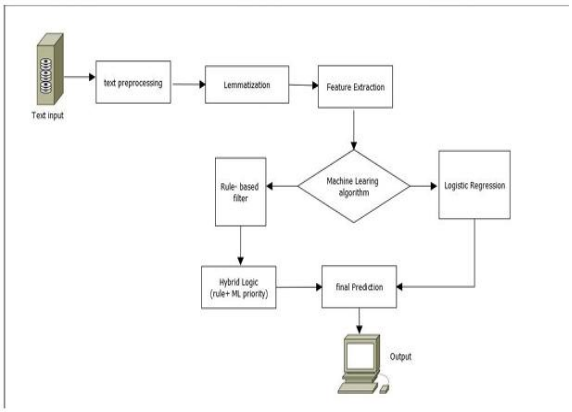

Fig. 1 System Architecture

System Architecture Description:

1. Input Layer:

Users provide input data, such as job descriptions, company names, and email attachments. The system also fetches live job postings via web scraping APIs or datasets like EMSCAD.

2. Data Preprocessing Module:

This layer handles missing data, duplicate removal, normalization, and text cleaning (stop word removal, tokenization, lemmatization).

3.Feature Selection and Extraction Module:

Important fields such as job title, company logo, education, experience, and salary are extracted. Irrelevant attributes are discarded to improve performance.

4. Model Selection Module:

The system allows users to select one of four algorithms — MLP, Passive Aggressive, Gradient Boosting, or KNN — for prediction.

5. Prediction and Trust Score Generation:

The selected model classifies the post as Fake or Real and assigns a Trust Score (20%, 60%, or 95%).

6. Explanation and Reporting Module:

This layer generates a report explaining why a post is fake (e.g., missing logo, mismatched company details).

7. Visualization Layer:

Graphical charts display trends such as ―Fake vs Real Posts Over Time‖ and ―Most Common Fraudulent Categories.‖

8. Output Layer:

The results, including classification, trust score, and explanation report, are presented to the user via a web interface.

Data Preprocessing Module (Detailed)

The data preprocessing phase plays a crucial role in preparing high-quality input for the model. Since online recruitment data often contain noisy and unstructured text, this step ensures consistency and reliability before prediction.

Steps in Preprocessing:

1. Data Cleaning:

Removes duplicate entries, null values, and irrelevant fields like timestamps or unrelated URLs.

2. Text Normalization:

Converts all text into lowercase and eliminates special symbols or punctuation to maintain uniformity.

3.Tokenization:

Splits sentences into individual words (tokens) for easier feature extraction and analysis.

4.Stop Word Removal:

Common English words (like and, the, of) are removed as they add no predictive value.

5. Lemmatization/Stemming:

Reduces words to their base forms (e.g., hiring → hire), improving model generalization.

6.Encoding and Label Transformation:

Converts categorical data (like job type, industry) into numerical form using one-hot encoding or label encoding.

7. Normalization and Scaling:

Applies Min-Max or Z-score normalization to balance feature values, ensuring fair contribution during model training.

VI. FUTURE WORK

In the future, this system can be enhanced by integrating deep learning models such as BERT, LSTM, or Transformers for more accurate text understanding and contextual analysis of job postings. Real-time API integration with major job portals like LinkedIn or Indeed can further automate data collection and detection. Additionally, incorporating blockchain-based job verification, multilingual support, and adaptive learning mechanisms will improve system transparency, scalability, and global applicability. The inclusion of a browser extension or mobile app could also enable users to instantly verify job authenticity while browsing online recruitment platforms.

VII. CONCLUSION

The proposed Online Recruitment Fraud Detection System offers an intelligent, real-time, and explainable solution to identify fake job postings using multiple machine learning models. By integrating efficient preprocessing, key feature selection, trust score generation, and detailed reasoning reports, the system enhances detection accuracy and transparency. It not only safeguards job seekers from fraudulent recruiters but also supports organizations in maintaining trust within digital hiring platforms.This approach demonstrates the potential of AI-driven methods in strengthening cybersecurity and ensuring safer online recruitment environments.

REFERENCES

[1] S. Kumar and R. Gupta, ―Detection of Online Recruitment Frauds Using Machine Learning

Techniques,‖ International Journal of Computer Applications, vol. 182, no. 48, pp. 1–6, 2020.

[2] N. Singh and M. Sharma, ―Fake Job Posting Detection Using Data Mining Techniques,‖ IEEE Conference on Computational Intelligence and Communication Networks, 2021.

[3] P. L. Li, ―Cybercrime in Online Recruitment Platforms: Emerging Threats and Countermeasures,‖ Journal of Cybersecurity Research, vol. 5, no. 2, pp. 78–88, 2019.

[4] A. Khan and D. Patel, ―Machine Learning Approaches for Detecting Online Scams,‖ Procedia Computer Science, vol. 198, pp. 120–127, 2022.

[5] B. K. Soni, ―Naive Bayes Based Classification for Job Scam Detection,‖ International Research Journal of Engineering and Technology, vol. 9, no. 3, pp. 245–251, 2022.

[6] R. Dash et al., ―Logistic Regression Approach for Detecting Fake Job Postings,‖ IEEE Access, vol. 8, pp. 98763–98772, 2020. 9

[7] H. Tan and J. Xu, ―A Survey on Online Recruitment Fraud Detection: Challenges and Trends,‖ IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 4, pp. 550–562, 2022.

[8] Y. Zhao and S. Wang, ―Effect of Feature Selection on Fake Job Detection Accuracy,‖ Applied Artificial Intelligence, vol. 36, no. 12, pp. 941–954, 2022.

[9] J. Chen and T. Zhou, ―Explainable AI in Recruitment Fraud Detection,‖ ACM Transactions on Intelligent Systems and Technology, vol. 13, no. 3, pp. 1–18, 2023.

[10] S. Nair and M. Joseph, ―Comparative Analysis of Machine Learning Models for Online Scam Detection,‖ IEEE International Conference on Data Science and Engineering, 2021.

[11] K. Patel and A. Bhattacharya, ―Trust-Based Scoring Mechanism for Online Fraud Detection,‖ Journal of Information Security and Applications, vol. 64, 2023.

[12] L. Reddy and V. Bansal, ―Hybrid ML Models for Online Job Post Authentication,‖ International Journal of Intelligent Systems, vol. 38, pp. 1209–1225, 2023.

[13] P. Singh and S. Sharma, ―Text Preprocessing in NLP-Based Fraud Detection,‖ International Journal of Computational Linguistics, vol. 12, no. 4, pp. 305–314, 2021.

[14] T. Mehta et al., ―Feature Engineering for Detecting Deceptive Job Advertisements,‖ IEEE Transactions on Computational Social Systems, vol. 9, no. 2, pp. 465–474, 2022.

[15] F. Ali and D. Chen, ―Linguistic Cues and Sentiment Analysis in Scam Detection,‖ Knowledge-Based Systems, vol. 235, 2022.

[16] H. Zhang, ―Explainable AI for Transparency in Cyber Fraud Detection,‖ IEEE Intelligent Systems, vol. 37, no. 6, pp. 32–41, 2022.

[17] R. Dey, ―Ethical AI in Fraud Detection Systems: Balancing Accuracy and Interpretability,‖ AI Ethics Journal, vol. 4, no. 1, pp. 15–27, 2023.

[18] M. Roy and K. Jain, ―Statistical Trend Analysis of Fake Job Posts in Online Portals,‖ Procedia Computer Science, vol. 217, pp. 34–42, 2023.

[19] A. Thomas and B. Narayan, ―Real-Time Fake Job Detection Using Web Crawling and ML,‖ IEEE International Conference on Smart Computing and Communications, 2023.

[20] J. Verma and S. Das, ―Multi-Model Machine Learning Approach for Fake Recruitment Post Detection,‖ IEEE Access, vol. 11, pp. 13475–13486, 2023.

[21] M.S.M. Anitha, Y. Naga Malleswarao, and P. Ajay, ―Online Recruitment Fraud Detection Using DL Approach,‖ Journal of Engineering Sciences, 2025.

[22] B. Alghamdi and F. Alharby, ―An Intelligent Model for Online Recruitment Fraud Detection,‖ Journal of Information Security, 2019.

[23] S. Mahbub, E. Pardede, and A.S.M. Kayes, ―Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries,‖ IEEE Access, 2022.

[24] P. Vara Prasad, N. L. Sravya, M. S. Kavya, V. Kavitha, and Shaik, ―Online Recruitment Fraud Detection Using Deep Learning,‖ International Journal of Progressive Research in Engineering Management and Science (IJPREMS), 2025.

[25] S. Dutta and S.K. Bandyopadhyay, ―Fake Job Recruitment Detection Using Machine Learning Approach,‖ International Journal of Engineering Trends and Technology (IJETT), 2020.

[26] K. Taneja, J. Vashishtha, and S. Ratnoo, ―Fraud-BERT: Transformer Based Context

Aware Online Recruitment Fraud Detection,‖ Discover Computing, 2025. of the IEEE/CVF CVPR 2023, pp. 18121-18131, June 2023.