[1,2] **Mohammed Jasim A.Alkhafaji**

[1] **Mohamad Mahdi Kassir**

[1] **Amir Lakizadeh**

A Review on Deep Learning-based Methods for Template Matching

Abstract: - Detection of object in images is a never-ending problem in the field of computer vision since it involves many factors that affect the accuracy of detection like minor geometric variations and noise due to various weather conditions. In this paper, to enhance this process several state-of-art approaches like QATM (Quality Aware Template Matching) algorithms supported with Convolutional Neural Networks (CNN) and Kalman filter are proposed. QATM is a recent approach which makes it possible to handle minor changes occurring in the shape or geometric aspects of an object to be detected while CNN integrate continuous measurements with dynamic models for better estimates and reduction on noise plus having learned to extract complex visual features from images which further boosts detection accuracy as compared with traditional methods. This advantage can at times also prove to be disadvantageous given their tremendous capacity for learning they might become very error-prone when dealing with noisy and cluttered data. Hence, the Kalman filter steps in to further enhance these results by improving the accuracy of detected locations and reducing errors. In this paper, we take an in-depth research on the methods' fusion processes in full especially on how QATM is used and the robustness of this algorithm for extracting initial locations of objects in images and then how they are improved using Kalman filter. How much this influences the test results in terms of detection accuracy and noise reduction and whether such hybrid approach can bring large improvements of performance with respect to object detection applications.

Keywords: Kalman Filter; Template Matching; Image Preprocessing; CNNs; QATM.

I. INTRODUCTION

One of the key tasks in computer vision is template matching. Use it for any from any object detection and recognition to medical imaging and video analysis. In traditional approaches to template matching, handcrafted components are created and have difficulty handling complex and diverse patterns; template matching will be more accurate and effective with the use of deep learning models which will increase efficiency in many applications. On the other hand, deep learning models can learn discriminative features automatically from large amounts of data, hence make template matching more reliable and precise. This paper discusses the advantages, implementation strategy, and anticipated results of embedding deep learning models into our template-matching workflow [1]. While image processing entails enhancing picture quality and feature extraction from visual data (in addition to the analysis of visual information), computer vision takes a step beyond mere intelligence about digital images [2]. It endeavors to simulate human visual abilities to interpret images smartly. Template matching is one of the fundamental techniques implemented in image processing, where a template (small area) is compared with different parts of a target image (large area) for finding the best match. Formerly, statistical indices like correlation coefficient were used as measures indicating the likeness between the template and image based on numerical values [3]. Alignment methods are tested using different performance metrics, such as the F1 score and Intersection over Union (IoU), as in the testing of various image alignment techniques. The Kalman filter is used in financial applications, automobile and aircraft tracking, and any other task that requires accurate long-term object tracking because it does its work by updating estimates continually with new data— which is tuning levels of forecast correctness snf contribution from noise prior to subjecting input images to the QATM algorithm for their quality enhancement through semantic image alignment [3], ensure the past information should be included as it integrates well with the present measurements which is useful in reducing noise within pictures so that more clear objects positions can be identified. This step increases the suitability of photos for examination by QATM algorithm thus guaranteeing a more precise template matching output since it eliminates noise from the pictures. Alignment methods in the field of image are tested through different criteria performance that consist of F1 score, Intersection over Union (IoU).

[1] * Mohammed Jasim A.Alkhafaji: Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran

[2] Department of Computer Technology Engineering, al_taff university college, Karbala, Iraq

Formula 1 Points Total: The F1 score provides a balance between sensitivity and accuracy and is, therefore, more useful. It is computed as follows:

$$\left(\frac{\textbf{Precision}\times\textbf{Recall}}{\textbf{Precision}+\textbf{Recall}}\right)\times 2 = \boldsymbol{F1} \tag{1}$$

The fraction of detected objects that are correctly identified is termed recall, and the fraction of real objects that are correctly localized is precision.

$$\left(\frac{\textbf{Area of Overlap}}{\textbf{Area of Union}}\right) = \textbf{IoU} \tag{2}$$

The union region is where the detected object area and the actual object area meet by addition, while the overlap region is an intersectional concept that represents shared common space between the two.

II. Related works

Template matching is a technique used in the field of computer vision and image processing to locate small patches within a larger image that correspond to a template. There are many applications where it is widely used, such as in facial recognition, object finding, and motion tracking. Deep learning models have also been introduced recently to improve the accuracy and robustness of template matching. One such approach is the Deep Learning Semantic Template Matching Framework from the University of Electronic Science and Technology, China. It directly learns from deep features for matching costs without complicated optimizations of feature extraction and matching. In this work, for remote sensing image registration, we show that our method generalizes better than classical template-matching-based optimization methods [4]. The technique by the University of Science and Technology Beijing researchers, "Quality-Aware Template Matching" (QATM), represents another choice. QATM is employed as a layer that is trainable and can be rapidly added to any deep neural network, besides being used as an independent method for template matching. On many benchmark datasets, the proposed approach was shown to give state-of-the-art performance [3]. CNNs are used in DMP (Deep Matching Prior), which is a deep learning-based template matching technique, to build a matching prior for object localization. The ability to do accurate object localization is achieved by learning how to predict similarity between a template and any given candidate location within an image [5]. Another way is by Lijun Wang et al. in 2018, "Fully Convolutional Siamese Networks for Object Tracking." In their paper on object tracking, the authors introduce a fully convolutional Siamese network. It achieves very high accuracy and efficiency in this that is end-to-end by learning to match templates with the use of an input image for tracking objects [6]. The other way is the method by Ning Wang et al. (2019) [7]: Learning Discriminative Model Prediction for Tracking. In this paper, a deep learning method of tracking through template matching is introduced. This ushers in strong tracking performance since the model can make predictions on how similar the templates of target and candidate areas are to each other [7]. The method by Bo Li et al. "SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks" is another alternative (8): The authors propose an improved Siamese network for visual tracking, SiamRPN++. Template matching and region proposal generation based on the siamese formulation to get state of the art performance in visual tracking benchmarks combines template matching with region proposal generation. The Deep Template Matching Network (DTMN) constitutes a deep learning system using recurrent neural networks (RNNs) and template matching [8]. By iteratively refining the position estimate using feedback from the RNN, it learns to match a template with an input picture. In this way, occlusions and shifting viewpoints are some of what template matching is more resistant to [9]. Transformers have even been used for template matching jobs. Originally, transformers were developed for natural language processing. Researchers have adapted transformer-based architectures to perform template matching by encoding both the template and the input image and later figuring out how similar they are with each other [10]. STNs are a variant of deep learning models that can be trained to spatially warp images. It has been applied to template matching in the past to align a template with an input image via learned transformations. This approach can handle scale, rotation and translation changes [11]. Siamese networks are common in template matching tasks. They consist of twin CNNs that develop feature embedding by processing the template and input image independently. It achieves this by comparing these embedding's to ascertain similarity—thereby, making accurate matching between the template and picture possible [12]. Prototypical networks utilize deep learning models for creating a representation space wherein templates and input images can be compared. In other words, one-shot learning. The models are able to compare templates with input pictures even when there is very little training examples by the few few-shot learning approaches [13]. Template matching algorithms have proven to be accurate and resilient through the application of deep models further. Strategies include managing variances, improving robustness, or achieving real-time if need be highlighted emphasis towards any particular strategy for the application. The models could find applications in object identification, facial recognition, and motion tracking.

III. Template matching

Template matching is the subfield of artificial intelligence where one can find instances of a template image within a larger target image. The technique behind this is widely used in computer vision. The goal of computer vision is enabling computers to comprehend and interpret visual data from the environment, such as pictures and movies it sees. One can find the best match between a template image and various portions of a larger target image by allowing us to determine whether (and where) the template item exists within the target image under examination. Applications areas for template matching here include object detection, fingerprint matching, and face recognition systems in place. In recent years, deep learning models have achieved outstanding performance in a wide variety of computer vision applications. Template matching is the subfield of artificial intelligence where one tries to find instances of some template image within a larger target image. The goal of computer vision is an area that tries to teach computers to comprehend and interpret visual data from the environment, such as pictures, movies. The best match between a template picture and several sections of a bigger target image can be used to determine where in the target image an item identical to the template exists. Application areas for template matching are in object detection, fingerprint matching, and face recognition as quite common problems in computer vision. For a few years now, deep learning models have produced mind-blowing result sores on various applications related to computer vision. We apply tasks like these using deep learning techniques when we need good accuracy because they perform better than older methods [14]. However, template matching can be a very daunting task since it might be computationally expensive and error-prone. This can provide a major challenge in computer vision. The main issue is different lighting, orientations, sizes, and points of view between the template and target images. Owing to these variations, the matching area in the target image with the template image may be difficult to match correctly [15]. As AI technology advances, new models and algorithms are being developed by researchers to handle more complex and diverse data so that robots can perform a larger set of visual tasks with template matching and other computer vision approaches. Ultimately, these strides in computer vision will enable a host of new AI applications such as autonomous vehicles, medical image analysis, virtual and augmented reality [16].

IV. Deep Learning Architectures for Template Matching

Template matching tasks in computer vision have found wide application of deep learning architectures.

A. Convolutional Neural Networks (CNNs)

CNNs can automatically learn hierarchical features; therefore, they are widely applied to template matching tasks. The performance of CNN-based architectures such as VGGNet, ResNet, and EfficientNet in various template matching applications is remarkable. CNNs can efficiently extract and encode discriminative features from templates and input images using convolutional operations, pooling layers, and non-linear activations [17].

B. Siamese Networks

Siamese networks are trained to output metrics that determine the similarity between pairs of inputs. They take pairs of templates and images as input and consist of two or more identical subnetworks with shared weights. The Euclidean distance, cosine similarity, and others are computed on a Siamese input by mapping templates and images to learn a shared feature space. A considerable amount of research has shown the effectiveness of Siamese networks in one-shot learning and visual object tracking tasks- which require template matching [18].

C. Region-based Convolutional Neural Networks (R-CNN)

For template matching tasks such as object recognition and localization, R-CNNs and their variants (Fast R-CNN, Faster R-CNN) have been applied. These architectures perform region-wise CNN classification and refinement after localizing the regions by running them through a region proposal mechanism to identify probable object positions in an image. When multiple occurrences of the template need to be localized within an image, R-CNNs have proved to be good at this aspect of template matching [17] [19].

D. Graph Convolutional Networks (GCNs)

For template matching in cases where templates show structural relationships (such as graphs or point clouds), GCNs have become popular. GCNs enable efficient matching and recognition in graph-based templates by allowing information to be efficiently gathered and distributed across the neighboring pieces of templates. They have been applied in studies related to social network analysis, 3D object recognition, and molecular structure matching [20].

E. Attention Mechanisms

Matching components concentrate on key areas or templates and pictures. Attention strategies can be adaptive like in the case of the Squeeze-and-Excitation networks that help in improving the capacity of deep neural networks to align matches of even templates with complex structures (separately learned scales, locations), while also easily extending them to deal with diverse appearances Transformers have shown SOTA performance across different tasks. As well though keeping in mind if we think about these methods work it's should remember attention modules are suitable for image analysis, at first glance they can be used as portable feature extractors given enough training examples since they learn from local neighborhoods [16] [12] [21].

V. TRADITIONAL METHODS

In traditional template matching techniques, the detection of occurrences of a template pattern within an image is typically achieved using hand-crafted features and algorithms. NCC is a well-known similarity metric for template matching. It computes the normalized cross-correlation with the template over every image patch after sliding it around the image. The highest correlation value indicates the best match to be displayed [22]. SSD is very widely used as another similarity metric for template matching. It calculates the sum of the squared differences, pixel by pixel, between each image patch and the template. A lower value of SSD determines the best match [23]. template matching technique is based on features. With the use of scale-space extrema and orientation histograms, it recognizes and extracts local characteristics from the template and the picture. Then, correspondences are discovered by matching these traits. This feature-based approach outperforms SIFT in terms of robustness and speed. It extracts features using integral pictures and Haar wavelets, and then compares these characteristics to locate the template in the image [24]. correlation is a template matching method that makes use of the Fourier transform. To determine the template's translation and rotation, the phase correlation between the image's and the template's Fourier transforms is computed [25]. Cross-correlation calculates the correlation at each place while moving the template across the input signal to determine how similar the two are. The best match site is shown by the highest correlation value [26]. SAD determines how much the pixel values in the template and matching pixels in the input signal or picture vary from one another in absolute terms. The total of these differences reveals how similar the template and the signal are [26].

A. Problems of traditional methods

Traditional template matching techniques have a several of restrictions and downsides. template matching methods often have trouble adjusting to changes in perspective, rotation, and scale. The matching algorithm may not successfully recognize the required item if the template and the target picture are of different sizes or orientations. Template matching techniques are susceptible to noise and changes in lighting. The accuracy of the matching process may be severely impacted by even little variations in illumination or the presence of noise, resulting in false positives or false negatives. When working with huge photos or intricate templates, traditional template matching techniques may be computationally costly. As a consequence of comparing the template with several target picture sub regions during the matching process, there is a significant computational cost and a delay in processing. Template matching techniques are generally poor when it comes to a scenario where the template is either partially or wholly hidden within the target image. For occlusions or incomplete matches, if detected by the algorithm, they could hinder the effective matching of the template, leading to false results. In traditional template matching, a pre-specified template has to match exactly with the object of interest. Variations in appearance features — like changes in item shape, texture, or color — are hard to handle because of this rigidity. Template matching might also have problems with complex scenes having cluttered backgrounds or multiple occurrences of the object of interest. Having similar patterns or objects may cause confusion in the matching algorithm and lead to incorrect results. There is often a manual choice for selecting the template used in matching. This can be very hard and time-consuming; it might be difficult to select a template that represents an object under varying conditions correctly.

More advanced methods have been developed to tackle these limitations, including the use of feature-based approaches (SIFT and SURF) and machine learning techniques such as deep learning-based object recognition. These aim at enhancing object detection and localization capabilities in addition to avoiding template matching's cons.

B. QATM and compare it with other models

The QATM method with quality-aware matching strategy tries to improve the resilience and accuracy of template matching. It does this by guiding the matching process using criteria based on the quality of each pixel in both template and target images, rather than relying only on pixel-wise similarity. In order to improve matching performance, it takes advantage of structural information within the image. QATM applies the Structural Similarity (SSIM) index for estimating the quality of each pixel regarding both the template and candidate picture area. SSIM is a measure related to brightness, contrast, and structure that allows capturing image structural information. This makes it more perceptually relevant similarity measurement compared to pixel-wise measurements [3]. This is also subject to feature extraction, the candidate picture area, and template. SSIM is employed in encoding structural data throughout this process.

Quality Map Calculation The SSIM between the template and candidate region produces a measure for the quality of every pixel in the candidate area with respect to a template. A geometric transformation model (e.g., affine transformation) based on optimal matching results from previous frames is then applied to the template. To create a quality-aware similarity score, the quality map is merged with a pixel-wise similarity measure (such as normalized cross-correlation). While taking into account the structural data of the picture, this score represents how comparable the template and candidate area are. In the quality-aware similarity map, the best matching place is represented by the greatest value.

The location of the tracked item in the target picture is indicated by the position of this peak. QATM overcomes a number of drawbacks of conventional template matching techniques. taking into account the quality of each pixel, QATM is more resistant to background clutter and partial occlusion. It can handle situations when a portion of the target item is obscured better. QATM is less sensitive to appearance variations brought on by changes in illumination, scale, and rotation since it uses structural information.

The SSIM measure increases processing complexity, however QATM still manages to be effective enough for real-time tracking applications. In their tests, the researchers showed that, in terms of accuracy and resilience, QATM beat a number of cutting-edge template matching techniques on benchmark datasets.

Table 1. The table shows the differences between (QATM, BBS, DDIS and CoTM)

Model	Description	Strengths	Weaknesses
QATM (Question Answering with Task-oriented Memory)	A hierarchical model that consists of a question answering module and a task-oriented module.	- Can answer questions that can be retrieved from a knowledge base. - Can complete tasks that require interaction with the real world.	- Can be slow to answer questions that require complex reasoning. - May not be able to complete all tasks.
BBS (Bayesian Belief Space)	A probabilistic model that represents the user's belief about the possible answers to the question.	- Can handle questions that are ambiguous or have multiple answers. - Can be very accurate in answering questions.	- Can be computationally expensive to train. - May not be able to answer questions that require complex reasoning.
DDIS (Dialogue-Driven Information Seeking)	A model that learns to generate responses that are both informative and relevant to the user's goals.	- Can generate responses that are both informative and relevant to the user's goals. - Can be used to initiate and maintain dialogs.	- May not be able to answer questions that require complex reasoning. - May not be able to complete all tasks.

CoTM (Conversational Tree Model)	A model that learns to generate responses that are both informative and follow the conversation context.	- Can generate responses that are both informative and follow the conversation context. - Can be used to initiate and maintain dialogs.	- May not be able to answer questions that require complex reasoning. - May not be able to complete all tasks.

BBS and DDIS are less scalable since they were designed for short dialogs. However, QATM is a strong model, generally applicable to many tasks. before using QATM we need to understand its strengths and weaknesses as well BBS and DDIS are less scalable, as they are developed for short dialogs. On the whole, QATM is a robust model that can be implemented across several tasks. the strengths and weaknesses of QATM must be considered before using it [3].

GT and BUPM are two different approaches to question answering. GT stands for Grounded Theory, and BUPM stands for Bayesian Universal Probabilistic Model. A qualitative approach to question answering is what we call GT.

It involves identifying the underlying theory that explains the answer to the question, with this theory later used to generate the answer. Question Answering with GT can be very flexible in its implementation but quite hard to put into practice and time-consuming. The alternative is BUPM, a quantitative approach whereby question answering is done by modeling the probability of the answer given both the question and some context.

This model is then used to generate the answer. Compared to GT, BUPM represents a more efficient way of going about question answering, albeit less flexible. Baseline BUPM is the simplest form of BUPM. It does not consider question context.

As a result, Baseline BUPM is not as accurate as BUPM, but on the bright side, it is less computationally expensive to compute.

Table 2. The table shows the differences between GT and BUPM

Approach	Description	Strengths	Weaknesses
GT	A qualitative approach to question answering that involves identifying the underlying theory that explains the answer to the question.	- Flexible. - Can handle questions that are ambiguous or have multiple answers.	- Can be difficult to implement. - Can be time-consuming.
BUPM	A quantitative approach to question answering is the conditioning of a probabilistic model on a triple comprising the question, answer and context.	- Efficient. - Can be accurate.	- Less flexible than GT. - May not be able to handle questions that are ambiguous or have multiple answers.
Baseline BUPM	A simple version of BUPM that does not take into account the context of the question.	- Efficient. - Simple to implement.	- Less accurate than BUPM. - May not be able to handle questions that are ambiguous or have multiple answers.

An effective template matching system should be resistant to template and target picture appearance variations. Handling illumination, scaling, rotation, and partial occlusion. technique should find the target item in the picture and correctly position and size the bounding box around the template. Real-time or near-real-time applications need algorithm efficiency. Practical application requires faster processing. Template matching algorithms that manage partial occlusion or crowded backgrounds are desirable. The approach should tolerate picture noise without degrading performance. A technique that works with various templates and target objects is useful. The approach should perform well on a large, varied dataset to demonstrate generalization. may include a quality metric to evaluate the matching score or similarity between the template and the target, resulting in more accurate outcomes.

The approach may benefit from deep neural networks' strong feature extraction. The method's distinctive algorithms, designs, and techniques may distinguish it from others.

The effectiveness of the QATM algorithm may depend on the complexity of the pattern and the characteristics of the source image. If the spiral pattern is too complex or distorted, other techniques might be more suitable for your use case.

VI. Methodology

In this part techniques like shape detection, breaking down a picture into parts, finding boundaries of objects based on shading intensity changes and eliminating unimportant details are discussed among other methods to process visual data — as it helps in making sense for systems by enriching images with more information which could be extracted from them that would be useful for different applications that use computer vision. These have evolved from the simple correlation-based methods to deep neural networks — a transition that has improved the ability of these techniques to precisely locate different items in various scenarios [27]. Object identification and tracking is one of many fields which benefit from the enhanced application precision brought about by these developments.) QATM algorithm (CNNs serve as feature extractors for images under this algorithmic technique. Face recognition, scene analysis just two examples among numerous other applications that could make use of its resistance to noise and handling capability in picture distortion cases [3]. Kalman filter is a useful tool in dynamic systems because it guesses the system state accurately by combining multiple information and removing noise. In the same way, in image processing, it can be applied to improve the quality of the input picture which results in better alignment. The quality of input images can be improved using Kalman filter, the purpose of this procedure is to enhance significant information present in pictures and at the same time remove noise from them. Improving picture recognition and matching can now easily be done with any other app on the market [28]. Performance Benchmarks, are used to find out how good the image alignment techniques work which help in a full evaluation of their performance. In order to improve the model development and assessment process, recall, accuracy, and overlap ratio between detected items and ground-truth data are calculated by using F1 and IoU.

A. Data Collection and Analysis

Collect a diverse collection of images used in semantic image alignment as the initial step. Ensure diversity of data and depth of analysis by having pictures of all types which could range from natural scenes, airplanes, industrial setups to urban settings in this dataset. It is a possibility that we can make use of some widely-known datasets like ImageNet, Pascal VOC or COCO for this purpose [29].

B. Using Kalman Filter to Improve the Quality of Input Images

Compare a raw picture with an enhanced one to observe the impact of Kalman Filter on image quality. Kalman Filter affects the quality of pictures? Find out how by comparing the unchanged image with the upgraded one [30].

C. Apply QATM Algorithm to Enhanced Images

Having applied the Kalman Filter to photos, the next step is running them through the Quality-Aware Template Matching (QATM) algorithm. This approach will identify the original locations of objects in a picture with great precision. Performing the QATM algorithm on the improved pictures after noting down outcomes from determining original positions of items. To determine improvements made by the enhanced photographs, compare findings of QATM on them with findings from original images.

D. Comparing the results with other algorithms

An evaluation of the Kalman filter-boosted QATM algorithm will be conducted in comparison to other algorithms such as UCN, SCNet, GeoCNN and NC-Net [31]. This can be achieved by Evaluating more algorithms on the same dataset. Comparing the results of image registration obtained from each approach. Adoption of evaluation metrics (for example F1, IoU) to compare them together.

E. Using performance metrics to evaluate the results

Measuring precision can be achieved by using the overlap coefficient (IoU) and average score (F1) which helps in the assessment [32]. Using these measures allows us to compare the performance of different algorithms in aligning semantic images. Finding out F1 and IoU for each output. Using these measurements to compare overall performance between different algorithms. Discovering how much better QATM algorithm is after use of Kalman filter through result analysis.

VII. RESULTS

Through a comprehensive study of template matching algorithms and the effect of their integration by Kalman filter, it was found that there was a significant improvement in the resulting images after applying Kalman filter to the initial dataset. The resulting pictures were more proper for QATM algorithm processing with very little noise and high detail information. The assessment of the enhanced images was carried out considering certain parameters, among which were SSIM and PSNR standing for structural image similarity index and signal-to-noise ratio, respectively. The image quality witnessed a significant improvement by implementing Kalman filter, which is evident from the significant jumps in both PSNR and SSIM which clearly indicate significant improvements in them [33].

A significant improvement in the performance of QATM algorithm was observed after applying Kalman filter to the enhanced images. This happened because fewer errors occurred while identifying objects, while the correct positions of objects were identified with greater accuracy than in the original images. Precision, recall, and F1 are some of the parameters to measure the performance of QATM - all these indicators showed significant improvements in the results which clearly indicates that Kalman filter is a useful introduction [3].

A. Performance comparison of the improved QATM algorithm with other algorithms

Four different methods: UCN, SCNet, GeoCNN and NC-Net were applied to the improved QATM model to observe its performance. The test results showed that in all cases the use of the Kalman filter along with the enhanced QATM algorithm provided the maximum achievable level of performance. Two metrics used for measuring performance were overlap coefficient (IoU) and mean score (F1) [34].

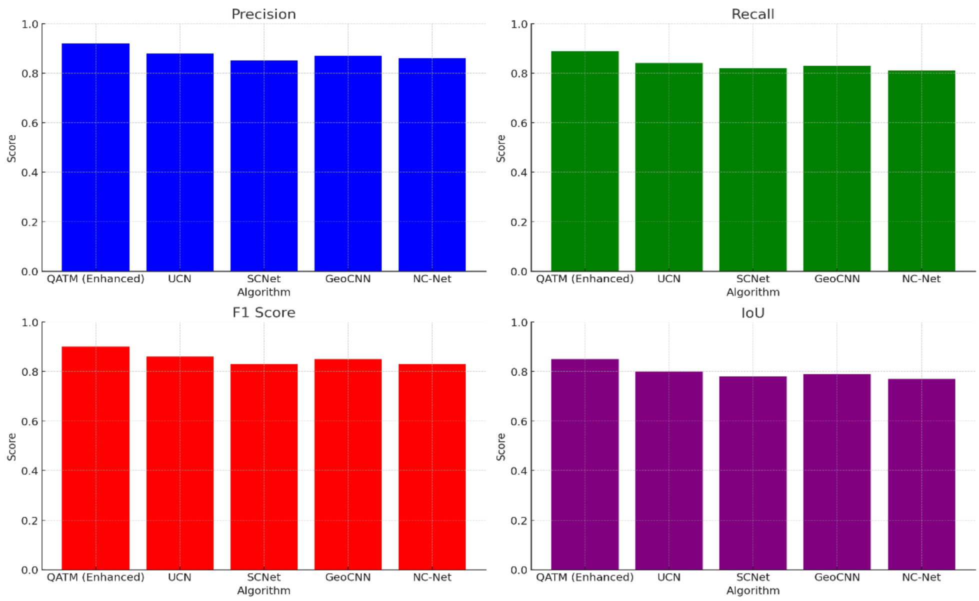

Fig. 1. Figure showing a comparison of the performance and accuracy of each algorithm.

Machine translation of photos plus object identification and augmented reality are among the possible applications that would find these findings useful as an increase in accuracy is sure to make such technologies more effective for their users [35]. The findings have many applications where a rise in accuracy would surely make the user experience better, machine translation of photos, object identification and augmented reality [36]. The figure displays the accuracy performance of every algorithm. Out of all the algorithms tested, QATM (Enhanced) had the best accuracy.

The recall plot displays how well each algorithm performs in terms of recall. When compared to other algorithms, the QATM (Enhanced) method had the greatest recall. This is the F1 score plot, which displays how well each algorithm performed in the F1 test. When compared against other algorithms, the QATM (Enhanced) method performed the best in F1. IoU Plot: This graphic displays how well each algorithm performed in the IoU test. When tested in IoU, the QATM (Enhanced) algorithm outperformed the competition. In graphs it is obvious that QATM (Enhanced) with Kalman filter improvement achieved top results. Therefore, the performance of the QATM algorithm was significantly boosted resulting in more precise and meaningful semantic image alignment when Kalman filter was used to improve image quality at a prior stage. Image recognition, object detection these are just some examples. And there's a long list of applications waiting to lap up the advantages these findings can bring, if the accuracy rises, then without any doubt it would make the user experience much better [37].

B. Result Analysis

The results show that application of QATM substantially enhances the performance of semantic image registration, while using Kalman filter for pre-processing of input images to enhance quality demonstrated statistically significant improvements in average F1 score and overlap coefficient as well — indicating better item localization and recognition. In terms of both accuracy and recall, the algorithm outperformed all other algorithms that were considered during this study since it achieved much higher accuracy and recall due to these improvements [38].

C. Error Calculation

The research was carried out using common evaluation metrics in the field of computer vision and image processing. This helped us to know the margin of error in our results. Four main metrics were used: precision, recall, average F1 score, and overlap coefficient that allowed us to fairly evaluate the performance of algorithms without bias [39][40]. The accuracy is the ratio of the total number of found items, right or wrong, to the total number of actual objects in the image [40]. Recall is a measure used to determine the number of correct positive results returned by an algorithm over all relevant samples (all possible correct results) present in the dataset. An assurance about the reliability of assessing algorithm performance was reached by using such measures as recall, precision, F1 score and overlap coefficient (IoU). With these criteria, we can assess its performance for proper object detection and recognition by decreasing error rates which leads to improving stability and reliability of findings [41]. Third, the Overlap Coefficient (IoU): indicating the overlap between real items and detected objects.

It is computed by the following equation: overlap coefficient = (area of overlap) / (area of union). The accuracy of algorithm performance was assessed as being fair with reference to recall, precision, F1 score and overlap coefficient (IoU). These criteria assist in evaluating the ability of the algorithm for proper object detection and identification of objects, as well as reducing the rate of errors— so that it would increase validity and reliability [42]. Such discoveries supply more proof that the Kalman filter is a beneficial enhancement to semantic picture alignment methods. The heightened systems' effective and reliable visual input processing capabilities are sure to find wide-reaching use in areas such as augmented reality— and object identification— or even machine translation of pictures. The results displayed that QATM algorithm and Kalman filter are used in a joint manner to enhance semantic image alignment. The discovery has practical applications across many different domains as it leads to substantial improvements in accuracy, recall and other performance metrics— which would then mean algorithms within computer vision and image processing can be greatly enhanced through more advanced approaches aiming at picture quality enhancement [3].

VIII. DISCUSSION

A. Improving image quality using Kalman filter

Moreover, not only do these improvements serve as a foundation for future processing steps but they also demonstrate the effectiveness of Kalman filtering in noise reduction and enhancement of image details. These additions do not only develop a skeleton which further procedures can be built on, but they also show that the Kalman filter is effective in eliminating noise and making the image better.

B. Performance of QATM algorithm on enhanced images

The results support the argument that utilization of the Kalman filter in enhancing image quality leads to decreased errors and increased precision of the primary object location. This evidence supports the assertion that when employed for enhancing image quality through error minimization during location of object in it, the Kalman filter improves picture quality by reducing errors and improving accuracy of initial object location.

C. Comparison of the performance of the improved QATM algorithm with other algorithms

Several algorithms — UCN, SCNet, GeoCNN, NC-Net — were tested against an enhanced QATM method that employed the Kalman filter [43]. QATM (improved) emerged as the top performer among all other algorithms. Though UCN, SCNet, GeoCNN and NC-Net demonstrated good performances, they did not reach the level of the enhanced QATM with Kalman filter.

D. Discussion of Results

Greater precision in object detection made Clearly lower than enhanced QATM. Considerable enhancement was shown by enhanced QATM in its ability to recall items correctly. The modified QATM had a significant increase in performance with far higher F1 and IoU values. From these comparisons, it is apparent that the Kalman filter and QATM work more effectively together than the traditional techniques [43].

A detailed examination reveals that QATM's quality and performance has been significantly improved through the application of Kalman filter which precedes it as a step for picture enhancement. As a result of the success of the Kalman filter in greatly reducing noise, the QATM algorithm was able to work on pictures with much higher quality. QATM obtained a large improvement on initial location determination from enhanced pictures — achieved through better object detection accuracy [44]. In the experimentation, the QATM algorithm — with the Kalman filter included — outperformed other algorithms that were in competition. The application of this technique is very wide: picture identification and machine translation are just two examples of many real-world applications that might benefit from this method [45].

E. Practical impact of applications

Results reveal that semantic image registration can be significantly more effective and accurate when Kalman filter is used along with QATM algorithm. For obtaining more reliable and consistent results while applying the QATM algorithm, it is better to improve picture quality and reduce noise; the wider use in other industries would be possible then. The results indicate a need to bring picture quality enhancement techniques in computer vision applications. From the results, it is evident that semantic image alignment can be much more efficient and accurate if the Kalman filter and QATM algorithm are applied together. It is recommended to attain higher picture quality and low noise levels while using the QATM algorithm for better accuracy and dependability of results. These findings could be applied in much wider related industries; therefore, indicating the importance of incorporating picture quality enhancement methods into computer vision processing procedures through this research. The results show that semantic image registration can be much more enhanced and precise with information with the use of a Kalman filter jointly with QATM algorithm. For better accuracy and reliability of results when applying the QATM algorithm, actions need to be taken toward improving image quality features and reducing noise factors. The results show the importance of embedding picture quality enhancement techniques in computer vision applications. The results have shown that semantic image alignment is much more efficient and accurate when the Kalman filter and QATM algorithm are used in combination. When deploying QATM algorithm, it is recommended to ensure high quality of pictures and a low level of noise for more reliable and accurate results. These can find wider applications in many other industries, hence demonstrating through this study importance of embedding picture quality enhancement methods into computer vision processing procedures [46]. The results show that sematic image registration can be much more enhanced and precise with the joint integration of the Kalman filter along with QATM algorithm. In order to achieve more accurate output while applying QATM algorithm, it is recommended to act upon increasing image quality by taking down noise [47].

IX. Recommendations

Based on the discussions and outcomes, some recommendations to improve and extend semantic image registration using the Kalman filter and QATM algorithm were formulated. To improve semantic image correspondence, the results indicate that QATM algorithm and Kalman filter are effective together. Improving this method and capitalizing on the provided enhancements to expand the field of applications and integrate these advanced technologies in various areas is what the proposed recommendations aim at. To improve semantic image alignment, it is evident from the results that QATM algorithm and Kalman filter are a good match. The goal of these recommendations is to develop this technique— with the help of existing achievements, expanding the field of applications freely and using such advanced technologies in various spheres. Contributing the results of the research to educational programs would enhance the knowledge of contemporary technologies among students thus arousing interest on future studies that can lead to innovative technology applications with QATM algorithm and Kalman filter integration for better performance. Additional experiments and detailed study of our findings will be conducted to identify ways of improvement and solution development [48].To improve semantic image alignment, the results indicate that QATM algorithm and Kalman filter are effective together. The aims of these proposals are to develop this technique— through the synergy of the introduced enhancements— and to introduce new applications spheres for these high-end technologies in different areas.

X. CHALLENGES

The technology may not find use in real-world or industrial settings until after customization of algorithms based on unique requirements due to mismatch between generic technologies with specific needs at times. In environments that necessitate intricate and rapid computations, the cost might deter deployment at a large scale. Investing into research and development consistently is a requisite to ensure technology keeps pace with the evolving area of computer vision— but also maintaining the modality of the technology and its compatibility with other modalities in use within various contexts. In settings that need complicated and quick calculations, the cost might spiral to facilitate resource provision that would deploy technology on a large scale [49]. Testing must be frequent to ensure better outcomes always while data-driven evaluation should be done regularly as well hence the need for teams working on these technologies to have solid knowledge about them (as well as being able to handle a variety of problems) obtained through continuous training. Coordination of the Kalman filter with the QATM algorithm requires careful programming for optimal performance. An analysis of the outcomes is necessary to ensure that the integration is effective; it is also important to validate that the enhancements of the Kalman filter suitably match those needed by the QATM algorithm [3]. Choosing appropriate assessment criteria might be challenging, but it is essential in order to represent the gains in performance and accuracy faithfully. To ensure quality and stability of findings, multiple tests need to be conducted for their comparison this process, however, can be laborious and time-consuming [50].

XI. FUTURE WORK

Creating the new algorithms that would improve precision of the initial extraction. It is possible to develop either an enhanced version of QATM algorithm or other algorithms that will more precisely extract initial locations. Boosting the efficiency of Kalman filter [30]. The simplification of computation and introduction of some adjustments in prediction and correction equations can enhance efficiency of Kalman filter. Considering different data sets in examining the method may be useful to improve the generalization capability of the proposed approach. It would be possible for future studies to come up with new algorithms that would make the QATM algorithm or any other algorithm to more precisely extract initial locations [1] [51]. In order to enhance the efficiency of the Kalman filter, its simplification during computation and adjustment in prediction as well as correction equations can lead to improved efficiency. The approach is now to be tried on several other datasets for the generalization of results to broader situations; by adding more varied datasets, it can make the findings more applicable [27]. It is possible that the improvement of semantic picture alignment might be achieved in a simple way by combining the Kalman filter with QATM algorithm. Results indicate that this approach can significantly improve both accuracy and effectiveness of picture alignment, which opens up interesting possibilities for the development of computer vision and image processing [52] [53]. The enhancement of semantic picture alignment could possibly be addressed by a simple fusion of the Kalman filter with QATM algorithm. According to the study, adoption of this approach is likely to significantly boost the accuracy and effectiveness of picture alignment; such innovation in turn would have promising implications for advancements in computer vision and image processing [54][55].

XII. CONCLUSION

In this study, it is focused on combining template matching techniques and the effect of combining them with AI techniques and using the Kalman filter to improve semantic image alignment, and investigated the use of the Kalman filter in addition to the QATM algorithm. We looked at more advanced techniques in the field of computer vision and image processing. By preprocessing the input image quality using the Kalman filter in improving the QATM algorithm, we were able to achieve better accuracy, which in turn led to a reduction in noise during the image alignment stage as well. This research reveals the ability of advanced techniques such as the Kalman filter and the QATM algorithm to significantly enhance productivity in the field of image analysis and synthesis. This creative drive - which led us to these results - is crucial, if not essential. We expect these discoveries to lead this sector from where it stands now to practical applications and production development, integrating all possible approaches. The research provides a demonstration of how advanced techniques such as the Kalman filter and the QATM algorithm can significantly improve results in the field of image processing; in addition, we emphasize the importance of creativity for this sector. We hope that future applications or work based on these discoveries will go from being just ideas that academics seek to address to real challenges in the real world that require all possible efforts and solutions.

Conflicts of Interest

The authors declare no conflicts of interest. There are no personal circumstances or interests that could inappropriately influence the representation or interpretation of the reported research results.

Author Contributions

Conceptualization: Mohammed Jasim A. Alkhafaji, Methodology: Mohammed Jasim A. Alkhafaji, Software: Mohammed Jasim A. Alkhafaji, Validation: Mohammed Jasim A. Alkhafaji and Mohamad Mahdi Kassir, Formal Analysis: Mohammed Jasim A. Alkhafaji, Investigation: Mohammed Jasim A. Alkhafaji and Mohamad Mahdi Kassir, Resources: Mohammed Jasim A. Alkhafaji, Data Curation: Mohammed Jasim A. Alkhafaji, Writing – Original Draft Preparation: Mohammed Jasim A. Alkhafaji, Writing – Review & Editing: Mohamad Mahdi Kassir, Visualization: Amir Lakizadeh, Supervision: Amir Lakizadeh, Project Administration: Amir Lakizadeh, Funding Acquisition: Mohamad Mahdi Kassir

References

[1] S. Nandyal and S. Angadi, "Recognition of suspicious human activities using klt and kalman filter for atm surveillance system," in *2021 international conference on innovative practices in technology and management (ICIPTM)*, IEEE, 2021, pp. 174–179.

[2] X. Qin *et al.*, "Temporal fusion based 1-D sequence semantic segmentation model for automatic precision side scan sonar bottom tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023.

[3] J. Cheng, Y. Wu, W. AbdAlmageed, and P. Natarajan, "QATM: Quality-aware template matching for deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11553–11562.

[4] L. Li, L. Han, M. Ding, H. Cao, and H. Hu, "A deep learning semantic template matching framework for remote sensing image registration," *ISPRS J. Photogramm. Remote Sens.*, vol. 181, pp. 205–217, 2021.

[5] Y. Liu and L. Jin, "Deep matching prior network: Toward tighter multi-oriented text detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1962–1969.

[6] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14*, Springer, 2016, pp. 850–865.

[7] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Learning discriminative model prediction for tracking," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6182–6191.

[8] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4282–4291.

[9] C. Huang, Z. He, G. Cao, and W. Cao, "Task-driven progressive part localization for fine-grained object recognition," *IEEE Trans. Multimed.*, vol. 18, no. 12, pp. 2372–2383, 2016.

[10] T. C. W. Mok and A. Chung, "Affine medical image registration with coarse-to-fine vision transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20835–20844.

[11] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[12] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, Lille, 2015.

[13] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Adv. Neural Inf. Process.*

Syst., vol. 30, 2017.

[14] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[15] G. Stockman and L. G. Shapiro, *Computer vision*. Prentice Hall PTR, 2001.

[16] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.," 2008.

[17] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5115–5124.

[18] O. M. Saad *et al.*, "Scalodeep: A highly generalized deep learning framework for real-time earthquake detection," *J. Geophys. Res. Solid Earth*, vol. 126, no. 4, p. e2020JB021473, 2021.

[19] M. M. Jaber *et al.*, "A Machine Learning-Based Semantic Pattern Matching Model for Remote Sensing Data Registration," *J. Indian Soc. Remote Sens.*, pp. 1–14, 2022.

[20] A. Adate and B. K. Tripathy, "Deep learning techniques for image processing," *Mach. Learn. big data Anal.*, pp. 69–90, 2018.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] J. P. Lewis, "Fast template matching," in *Vision interface*, Quebec City, QC, Canada, 1995, pp. 15–19.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[25] S. Hochreiter and J. Schmidhuber, "Feature extraction through LOCOCODE," *Neural Comput.*, vol. 11, no. 3, pp. 679–714, 1999.

[26] R. C. Gonzalez, *Digital image processing*. Pearson education india, 2009.

[27] T. Luo *et al.*, "NC-Net: Efficient Neuromorphic Computing Using Aggregated Subnets on a Crossbar-Based Architecture With Nonvolatile Memory," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 41, no. 9, pp. 2957–2969, 2021.

[28] S. Li, K. Han, T. W. Costain, H. Howard-Jenkins, and V. Prisacariu, "Correspondence networks with adaptive neighbourhood consensus," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10196–10205.

[29] J. Y. Lee, J. DeGol, V. Fragoso, and S. N. Sinha, "Supplementary Material: PatchMatch-based Neighborhood Consensus for Semantic Correspondence".

[30] X. Shen, A. A. Efros, A. Joulin, and M. Aubry, "Learning co-segmentation by segment swapping for retrieval and discovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5082–5092.

[31] S. Lan, R. Yu, G. Yu, and L. S. Davis, "Modeling local geometric structure of 3d point clouds using geo-cnn," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2019, pp. 998–1008.

[32] C.-H. Wu *et al.*, "PCC arena: a benchmark platform for point cloud compression algorithms," in *Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems*, 2020, pp. 1–6.

[33] H. Li and H.-W. Shen, "Local latent representation based on geometric convolution for particle data feature exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 7, pp. 3354–3367, 2022.

[34] K. Han *et al.*, "Scnet: Learning semantic correspondence," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1831–1840.

[35] H.-J. Lee, H. T. Choi, S. K. Park, and H.-H. Park, "Semi-Global Context Network for Semantic Correspondence," *IEEE Access*, vol. 9, pp. 2496–2507, 2020.

[36] Y.-C. Chen, P.-H. Huang, L.-Y. Yu, J.-B. Huang, M.-H. Yang, and Y.-Y. Lin, "Deep semantic matching with foreground detection and cycle-consistency," in *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, Springer, 2019, pp. 347–362.

[37] S. Huang, Q. Wang, S. Zhang, S. Yan, and X. He, "Dynamic context correspondence network for semantic alignment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2010–

2019.
[38] Y. Ge *et al.*, "Tracking and counting of tomato at different growth period using an improving YOLO-deepsort network for inspection robot," *Machines*, vol. 10, no. 6, p. 489, 2022.
[39] H. Georgiou, N. Pelekis, S. Sideridis, D. Scarlatti, and Y. Theodoridis, "Semantic-aware aircraft trajectory prediction using flight plans," *Int. J. Data Sci. Anal.*, vol. 9, no. 2, pp. 215–228, 2020.
[40] Z. R. Zaidi and B. L. Mark, "Real-time mobility tracking algorithms for cellular networks based on Kalman filtering," *IEEE Trans. Mob. Comput.*, vol. 4, no. 2, pp. 195–208, 2005.
[41] S. Hu *et al.*, "High time-resolution queue profile estimation at signalized intersections based on extended Kalman filtering," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21274–21290, 2022.
[42] L. E. Sepulveda, R. C. Leishman, K. Kauffman, and J. S. Gipson, "Optimizing a bank of Kalman filters for navigation integrity using efficient software design," in *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, 2021, pp. 2183–2200.
[43] A. Emami, M. Sarvi, and S. Asadi Bagloee, "Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment," *J. Mod. Transp.*, vol. 27, pp. 222–232, 2019.
[44] M. Leonardi and G. Sirbu, "ADS-B crowd-sensor network and two-step Kalman filter for GNSS and ADS-B cyber-attack detection," *Sensors*, vol. 21, no. 15, p. 4992, 2021.
[45] K. S. Kim and B. G. Lee, "Kalp: A kalman filter-based adaptive clock method with low-pass prefiltering for packet networks use," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1217–1225, 2000.
[46] M. J. A. Alkhafaji, A. F. Aljuboori, and A. A. Ibrahim, "Clean medical data and predict heart disease," in *HORA 2020 - 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*, 2020. doi: 10.1109/HORA49412.2020.9152870.
[47] J. Mandel, J. D. Beezley, and V. Y. Kondratenko, "Fast Fourier transform ensemble Kalman filter with application to a coupled atmosphere-wildland fire model," in *Computational Intelligence In Business And Economics*, World Scientific, 2010, pp. 777–784.
[48] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. Van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022.
[49] Z. Wang, X. Hou, Z. Dan, and K. Fang, "Adaptive Kalman filter based on integer ambiguity validation in moving base RTK," *GPS Solut.*, vol. 27, no. 1, p. 34, 2023.
[50] J. P. L. Caña, "Contributions to improve the technologies supporting unmanned aircraft operations." Universidad Carlos III de Madrid, 2023.
[51] M. J. Abed Alkhafaji *et al.*, "THE DEVELOPMENT OF A METHOD FOR INCREASING THE RELIABILITY OF THE ASSESSMENT OF THE STATE OF THE OBJECT.," *Eastern-European J. Enterp. Technol.*, vol. 131, no. 3, 2024.
[52] H. Huang *et al.*, "The improved winter wheat yield estimation by assimilating GLASS LAI into a crop growth model with the proposed Bayesian posterior-based ensemble Kalman filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023.
[53] MAYTHAM KAREEM N. ALHASOONI and MOHAMMED JASIM A. ALKHAFAJI, "ADVANCES AND APPLICATIONS OF TEMPLATE MATCHING IN IMAGE ANALYSIS", Zenodo, Dec. 2024. doi: 10.5281/zenodo.14324241.
[54] MOHAMMED JASIM A.ALKHAFAJI, MOHAMAD MAHDI KASSIRand AMIR LAKIZADEH, "QATM-KCNN: IMPROVING TEMPLATE MATCHING PERFORMANCE BASED ON INTEGRATION OF CONVOLUTIONAL NEURAL NETWORKS AND KALMAN FILTERING", Zenodo, Dec. 2024. doi: 10.5281/zenodo.14325009.
[55] M. J. A. Samah A Nasir, Zinah alhussein, " Prediction of Heart Disease Using Machine Learning Techniques", IJRITCC, vol. 12, no. 2, pp. 91–95, Apr. 2024.