Journal of Information Systems Engineering and Management
2025, 10(21s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Review on Deep Learning-based Methods for Template Matching: Advancing with QATM-Kalman-CNN Integration

Mohammed Jasim A. Alkhafaji [1,2], Mohamad Mahdi Kassir [1], Amir Lakizadeh [1]
[1] *Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran*
[2] *Department of Computer Technology Engineering, al_taff university college, Karbala, Iraq*

ARTICLE INFO

Received: 19 Dec 2024

Revised: 02 Feb 2025

Accepted: 16 Feb 2025

ABSTRACT

Introduction: Template matching algorithms have been widely used for identifying complex emissions in large datasets, offering a rapid initial assessment capability. However, their performance is often constrained by the requirement for prior training on target species and the high occurrence of false positives due to background noise interference. To enhance the efficiency and accuracy of these algorithms, we explore an improved approach leveraging deep learning techniques.

Objectives: This study aims to improve the robustness of template matching algorithms by integrating Quadrature Amplitude Time Modulation (QATM) as a preprocessing step for convolutional neural networks (CNNs). The primary goal is to optimize target species detection while mitigating false positives and enhancing performance in the presence of sophisticated interference.

Methods: The proposed method utilizes QATM to condition input data before feeding it into CNNs. CNNs, being one of the most fundamental and effective deep learning architectures, are employed for both training and prediction. By leveraging CNNs, we seek to automate the training process of target species detection within template matching algorithms, reducing dependence on manual intervention.

Results: Preliminary findings suggest that integrating QATM with CNNs enhances template matching performance, leading to a reduction in false positives and improved detection accuracy. The model demonstrates promising potential in handling complex signal interference, providing more reliable predictions.

Conclusions: This study highlights the feasibility of using CNNs in conjunction with QATM to enhance template matching performance. Future work will focus on refining this approach by conducting objective tests to fully automate target species training within template matching algorithms and further improve detection accuracy under challenging conditions.

Keywords: Template matching, QATM, CNN, Kalman filter, digital modulations, modulation classification algorithm

INTRODUCTION

This work studies the integration of QATM with convolutional neural networks (CNNs) through the Kalman filter (KF). A visual monitoring system is proposed, which consists of two parts, the feature extractor and the tracking model. It consists of two main modules, object and template matching. Object detection algorithms are used to estimate the sizes of the rectangular regions, denoted by D. Template matching algorithms are discussed while estimating the intrinsic similarity between the feature images of the window and the window. QATM is a fast template matching framework using CNN features, and can be set to enhance the results. The OPU is a bi-objective optimization algorithm for template matching using CNN features. The performance improvement of these three details is verified by experiments [1][2].

From an algorithmic perspective, template matching refers to the task of finding a template on an image or video. As various computer vision tasks are emerging rapidly, template matching algorithms have found broad potential in target recognition, video security, and pedestrian detection. The conventional template matching algorithms depend on the techniques in computer vision and AI systems. The normalized correlation coefficient and SIFT are the typical methods in 2D still images. The deep feature representation learned by CNN has dominated tasks such as object detection, object recognition, and semantic segmentation. These CNN features have the properties of scale invariant, rich pattern repertory, and geometric deformation invariance. Despite the existence of different CNN architectures, accurate feature representation is not a complete guarantee in template matching. A recent study uses the template samples to generate a codebook model and model the sample data for a hierarchical search. A diversified search and data queue are applied to improve the performance of QATM. Another OPU method utilizes the pretrained models of CNN to extract the features of the sample data, and edge-dependent windowing is applied to obtain the promising results in the object tracking system [3][4].

BACKGROUND AND MOTIVATION

Template matching is the problem of finding one or several items in an image that match in appearance to an input configuration or template. The aim is to improve the performance of template matching in terms of correct detections and false positives [5]. Template matching should inherently not depend on colors or not suddenly arise depending on the color of the background. The method is designed to handle slight deformations and occlusions without using any assumed shape prior. A comparison is made against existing methods on both the original version and the published more restricted version, with AUC at 82.37% and 62.83%, respectively. The version made available independently of the peer review featuring a new hashing methodology aroused suspicion on the part of the reviewers, and this was investigated. The concern related to the processing and analysis of pairwise distances, ultimately requiring a retraction. In this new version, Computation Section 2.2 was expanded to provide a description of the evaluated method. Section 2.3 was also added, describing the analysis on the parameter space. Despite peer reviews and suspicions in the language of the commentaries, the correctness and validity of the results codes were ultimately confirmed [2].

Since its introduction as a mathematical problem, the question of counting subgraphs of a given type in a random graph model has been an active area. The first work in this area is due to the seminal paper of Bollobás, who developed the basic theory of subgraph count statistics in random graphs. These results were generalised and applied to algorithmic problems by other researchers including McKay, Alon, Janson and Ruciński. The field has remained an active area of research, with the recent work of several researchers developing powerful new methods based on graph limits. These results have found applications to important problems related to the detection of surprising structure in machine learning and the analysis of social networks. However, all previous work remains focused on the study of dense subgraphs such as triangles or quadrangles. Until very recently, there were no results on the distribution of other types of subgraphs. An investigation of how these results can be adapted to the complete random intersection of indices of monomials is explored in the follow-up manuscript [6].

Gao and Spratling introduced a collection of python code which extends a template matching method in shape-biased CNN feature space. This code integrates a well-known and computationally efficient QATM codebase based on the RGB feature-space with the DIM algorithm to generate and process extra color channels (CIE-Lab, gradient-magnitude). The Gao and Spratling code also implement the Kalman filter to predict object trajectories. To assess the generality of these findings, a more extensive experimental evaluation is conducted on a collection of arbitrary objects from a real-world scenario and both imaged and computationally generated textures are employed [5]. This paper demonstrates that extending the BBS, DDIS and DIM algorithms to shape-biased CNN features with color enhancement leads to only a modest performance increase in some cases. However, applying the DIM-CNN with Kalman filter processing to object-based experiments results in a substantial performance improvement.

Gupta and Sintorn propose an efficient template-matching approach using VQ-NNFs for fast and robust high-resolution matching of 2D images in the wild. The proposed algorithm uses off-the-shelf domain-specific models applying them to the VGG19 network at different depths. This exercise swaps the traditional (RGB) image representation with an extension up to 5 channels that combines the standard color space with physically meaningful gradients aiding to diversify the spatio-color image content. The Gao and Spratling method works with 256-d templates, images are resized to 1024 px on the short side, and the original aspect ratio is preserved. The QATM-DIM algorithm competition is set to 10.000 iterations for the challenger and 8000 for the defender. All the object-centric experiments compared to the feature-space baselines were carried out using item shapes of single objects from the COCO2014 dataset. Rapid progress has been made in the field of computer vision to improve tracking performance and increase the range of situations in which a CNN can operate [7]. Several different methods have been developed that participate in object tracking latitude. Central to most approaches, however, is the requirement to detect the object in the first frame of a video. This presents a difficulty in uncontrolled environments when the object is concealed at its origin or when a scene change occurs.

RESEARCH OBJECTIVES

The proposed work consists of two objectives. The first objective is to explore the concept of QATM and prepare an implementation of QATM and a first draft of the associated paper. The second objective is to study how QATM can be used as a matching cost in CNNs and how CNNs can be integrated with QATM through the Kalman filter. After some reviewing, two potential approaches have been proposed which are relevant for this scene understanding workshop by [7]. With some adaptation, the described QATM algorithm and an approach involving the use of CNNs together with QATM as an intermediate feature representation could be combined, such that QATM is used as a matching cost in a CNN object detection framework, and the match tracklets are updated between frames using a Kalman filter.

Ultimately the output of this work would be a joint paper which could be submitted to a relevant computer vision conference, summarizing the described experiments and any additional relevant experiments relevant to the adaptations made by. This work, regarding the adaptation of the QATM algorithm and the evaluation of QATM as a matching cost in a CNN object detection framework, will be treated as the first objective, and the second objective as outlined above to be worked on should the first objective be completed in a timely manner [8].

SCOPE AND LIMITATIONS

The field of computer vision has always been dense and fertile ground for the intertwining of scientific intuition, creativity and technology innovation. Template matching plays a key role in the computer vision realm, and such importance is testi to the wealth of research papers published over the past decades. The successful application of the Quality-Aware Template Matching paradigm shows that domain knowledge-driven techniques that exploit the inner statistical structures of images can be more robust to occlusions, clutter and background clutter than deep learning techniques. Accordingly, we propose a study in the advancements in template matching algorithms and in the assessment of the integration of an existing image template matching paradigm (QATM) with convolutional neural networks (CNNs) through the Kalman filter. With an illustrative research experiment it is shown how non-rigid tracking is testing the limits of deep learning for near real-time scenarios, and quality concerned appearance models leveraging domain knowledge are more effective under a streaming scenario [9]. The results highlight the successfulness of deep-learning free QATM with respect to learnt feature-based methods in both static and transient occlusion cases, too. Due to ever increasing accuracies on the ImageNet benchmark, CNNs have gained unprecedented momentum in the computer vision research community. These networks have been successfully adapted to a wide range of computer vision tasks through off-the-shelf features or the fine tuning of the full network to a custom task. In the template matching area, there has been an upsurge of new algorithmic paradigms that can be used with off-the-shelf deep features. A new object appearance model for long-term tracking is proposed using the particle filter-guided QATM paradigm, based on the paradigm of BBS. This is demonstrated to be effective in complex scenarios with color clutter, partial occlusions, scale and perspective changes, and different poses. A GPU-accelerated and Explorer-like Python package is relea which embeds BBS, the particle filter, and QATM, able to directly execute experiments on a target test bed. Results of an extensive evaluation with respect to both state-of-the-art appearance models and deep-learning based techniques are given.

FUNDAMENTALS OF TEMPLATE MATCHING

Template matching is a fundamental problem in computer vision [7]. It measures the similarity between a fixed template and patches in an image. Results have many applications, such as object detection, image registration, and object tracking by matching the object's shape in a search image. A simple, yet effective, approach is to compare the template with a patch in the image using nearest-neighbour (NN) distance; the patch can be chosen as the same width and height as the template [5]. The NN-based comparison can be extended with techniques such as taking the average of distances of the top-K NNs. Aggregate comparison functions include computing the mean, median, and variance of the distance values within the neighbour and, consequently, form a histogram. Real-world data can exhibit problematic changes in scale, rotation, and articulation. Nyström extension is used as an ambiguation expansion in which the template patches ask for the best match from a pool of potential image regions. It can be circuited by fixing the template position and sliding it over the image. The most similar image patch is declared the match for the template. Classic template matching algorithms are illustrated in which possible implementations are provided, with an emphasis on the discrete, yet relevant, choices of hyper-parameters as well. There is the possibility to summarize such algorithms at different levels of abstraction and the key concepts behind them. Lead is provided to a novel algorithm which is the combination of QATM and the final observations about STR. Broad conclusions about the pros and cons of working in the STR domain are drawn. Every general claims and observed behaviours are supported by experimental results [10].

TRADITIONAL TEMPLATE MATCHING TECHNIQUES

To the human visual system, intelligence machines use pattern analysis on images. This can include object detection, recognition, scene interpretation, gesture and facial expressions reading, and motion analysis. Template matching is a simple, yet fundamental, concept that consumes data from the intelligence network, often used as the building blocks for more complex pattern analysis. A great variety of algorithms have been available for template matching in recent years. These algorithms have origins in the principal mathematics, analysis and imaging science disciplines. Among these traditional template matching algorithms, Sum of Absolute Differences, Sum of Squared Differences and Normalized Cross-Correlation are commonly used for image-based geometry-based pattern recognition tasks [11].

Instead of having to re-calculate all the results with numerous match patches, newer works have proposed algorithms that can update the results from the previous patch searching and eliminate the redundant calculations with the matched surfaces. A training knowledge model to be applied as a hall network is introduced. A comprehensive matching description metric is also presented. A neural network based on qualities is discussed. A pre-selection matching search zone is defined. A search picture based on targeted searching is described. Overall, the choice of patch review is a simple choice and the angle range for the view is not considered. The problem is that previous methods have not found appropriate image regions containing the source object, which may be only visible from specific viewpoints or far away from the camera. In contrast, we propose an approach that synthesizes a query template from an existing part or design model. To achieve the view representation, we employ a convolutional neural network architecture as an FCN. In order to take into account occlusions and clutter, the neighborhood of the object must be ignored in the matching process, indicated by bounding circles [12].

ADVANCED TEMPLATE MATCHING TECHNIQUES

Template matching is a fundamental problem in computer vision with applications in object detection, image registration and object tracking. The aim is to find all regions in an image that are visually similar to a template. The appearance of such regions are similar to the template, but the template might also represent a region in the image that is deformed. This can be due to e.g., perspective distortions in cluttered surroundings, occluding objects or non-rigid transformations.

Template matching is a challenging problem due to the high variability in images, the sensitivity to noise and the conflicting requirements of finding similar regions (i.e. low false positives on the template regions) and avoiding local minima due to the non-convexity of the cost function. Template matching algorithms search for such regions by comparing the template to the query image in a predefined feature space. The matches are often found by computing the distance between the template and a sliding window in the query, where distance functions like SSD or cross-correlation are commonly used. As a post-processing step, the similarity can be filtered with a smoothing function. In contrast to the studies described above, this work presents a study which integrates QATM with CNN by performing further filtering of the NCC similarity maps. The proposed method utilizes a Kalman filter to keep track of the locations of objects being searched for, so that NCC similarity values are only calculated in the vicinity of these positions. Used this way, the Kalman filter substantially reduces the number of similarity calculations needed, and it is shown that this does not significantly impede the object matching process, at least not under the noise conditions tested. By further filtering the similarity maps with a bilateral filter or dilating the binary object locations, accuracy is found to show a considerable improvement [13].

FEATURE-BASED METHODS

Template matching has long been a fundamental problem in computer vision. Matching a template representation to a larger image and finding its location receives a lot of attention in various fields, including object recognition, pose estimation, and image alignment. However, traditional template matching approaches often fail in real-world scenarios. In many applications, the appearance of the object and background changes, and there can be various transformations that need to be tolerated. Many template matching methods have been developed to address robustness to these changes, including feature-based methods, learning-based methods, and template disentanglement methods [14].

One of the greatest challenges of template matching has always been invariance to the seemingly infinite ways an appearance of an object can fail to match the template. Over the past decades, many techniques devoted to this issue and related problems have been developed. All template matching technologies may be divided into four main classes: area methods, profile methods, correlation methods, and spectral methods. All the mental and medical applications supporting graphical user interface have the great demand in development of the image processing functions including the area and correlation search for pattern recognition. Efforts to create a versatile function of the convolution type involving templates larger than the objects stimulated the development of many new technologies [7]. Among them, 'patented template matching' technology can be considered the evolution of the Template Transmission Technique that has demonstrated the most robust template matching throughout a wide range of applications. Generally, a search for a template on the image requiring the user to manually input the area with an example of this template and the sizes of patterns.

DEEP LEARNING-BASED METHODS

Template matching has been a basic procedure in computer vision for various applications, including image and video processing, object detection, feature alignment and tracking, pattern recognition, medical imaging, satellite image analysis, and iris matching in biometric systems. Traditional methods for template matching are pixel-based and invariant to either shape or texture information, thus they are sensitive to changes in features and background clutter. To overcome these drawbacks, many approaches have been developed for creating a new feature space using a CNN and integrating it with a classic template matching algorithm to compute a higher-level representation of the data in this space. Multiple fully connected layers with dropout and activations are trained to 'memorise' the invariant attribute of known images, while the required discriminative attributes rely mostly on the last convolutional layers. Quality-Aware Template Matching is a method that uses a pretrained CNN model as a feature extractor, where the highest feature matching score is achieved on CNN features. The goal is to optimize the same custom CNN architecture to perform well on the classic task of natural image classification while aiming to do this based on shape rather than texture cues [15].

A Kalman filter is proposed that processes template matching output to predict target state in consecutive frames. The method uses deep features from a shape-biased CNN, recently developed on template matching scores generated in the feature space, created to increase tolerance to changes in appearance and to be more discriminative with respect to shape-information. This feature space is general to the template and search image for template matching, making this method suitable for different tasks and scenarios. To the best knowledge, only five-dimensional representation is used for this purpose. The CNN-inspired feature space is built using multiple features already present from the template matching scores, exploiting the informative output of the state-of-the-art algorithm. It includes twelve-dimensional shape context of the template and search points in the complex feature space, six additional dimensions encoded using the feature position in the fixed closest triangle mesh, and a time series dimension for building the temporal relationship of tracks of the movement of particles in the flow. At each iteration, this process generates the state-space representation of the tracked feature. It was defined for a more complex task and environment where the natural shape similarity between the search and target object is reduced [16].

QUALITY-AWARE TEMPLATE MATCHING (QATM)

Much recent work in the field of computer vision has focused on devising advanced template matching algorithms. In the majority of real-world scenarios, however, the translation, rotation, scaling, or colour of the templates and/or the target images is unknown and does not follow a set motion model. Under such adverse conditions, the vast majority of common template matching algorithms are known to be inapplicable, which motivated the development of Quality-Aware Template Matching. The workflow and results of Quality-Aware Template Matching are outlined. The results of a new study are presented which show that, despite performing well individually, the outputs of Quality-Aware Template Matching and Quality-Aware Template Matching Refinement can lead to a further improvement in performance when used together. The suitability of integrating Quality-Aware Template Matching with convolutional neural networks for object tracking applications is investigated. The use of a Kalman-filter for tracking is outlined, and an analysis of the geometric conditions which must prevail between the bounding box of the template and its expected position in the target image is made. From this, a simple approach for extracting the estimated translation, rotation, and scale changes from the expected position of the template is proposed. By following the method described, the optimal tracking performance that can be achieved under the assumption that the templates are perfectly tracked is represented [17].

OVERVIEW AND PRINCIPLES

Motivation, Goals and Contributions - Since the advent of deep learning, it has been conflicting how such networks can be used as a computational model of biological brains, which are in significant contrast to deep learning. Throughout the years, neuronal algorithms have surged due to their biological plausibility and robustness. However, most current neuronal algorithms are implemented by homemade code rather than deep learning libraries due to their non-differentiable properties. In such circumstances, it is important to seek a gateway for bridging deep learning of the power to neuronal algorithms. The goal of this work is to represent a comprehensive framework to effectively integrate the quantized average templates matching algorithm with convolutional neural networks in a differentiable manner via the Kalman filter. Condensed results justify the superiority of the proposed framework over the counterparts [18].

Overview and Principles - Template matching is a simple but important problem in computer vision with applications in object detection, object tracking, and so on. Among the different template matching methods proposed, the quantized average templates matching algorithm is a simple but efficient method to encode the information in the MRFways, which is easy to implement and can achieve state-of-the-art performance. Challenge arises when the quantized average templates matching has high-resolution templates. Motivated by the pitfalls, this work integrates the quantized average templates matching with convolutional neural networks in a differentiable way. However, such an interface is non-trivial due to the discrete and non-differentiable properties of quantized average templates matching. To tackle this issue, this work leverages the smooth mechanism in the optimization procedure to form smooth quantized average templates matching and a differentiable Kalman filter is designed to differentiate the filtering process. The essential principles of the proposed framework are elaborated in this section [19].

APPLICATIONS AND BENEFITS

In order to illustrate the broad applicability and the benefits of the invention six different applications are studied: keypoint matching in images, tracker seeding in videos, deforming object segmentation where template matching is used to find an object region in a query frame given an initial object region in the template frame, ransac voting, stereo compensation, and optical flow. These applications are chosen to cover scenarios of both planar and non-planar template-camera motion and both rigid and non-rigid object motion [20].

Responding to co-invented advantages of a template matching algorithm, a study on QATM is proposed. Combining the invented QATM with deep convolutional neural networks and implementing a proposed QATM-CNN method is also aimed in this proposal. In this work, QATM filtering is designed cooperatively with CNN, and the output heatmaps of CNN is corrected using kinematical information of an existing or a modified object trajectory model by Kalman filtering. This study expects the feasibility of object detection in complex scenes with deformable objects by the integration of CNN with QATM and Kalman filter [21].

CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Template matching is a powerful method to localise target objects in a noisy environment. The Quickest All Template Matching method applies a bank of optimised filters with different sizes and orientations to each frame to detect a wide range of possible targets. Detections are classified and passed to a Kalman filter and a template matching test. A study was conducted on the integration of QATM with deep features and convolutional neural networks equipped with colour and shape–texture debiasing methods. A bank of optimised filters for deep features significantly reduces the search space and speeds up the algorithm. The QATM detections have an average overlap with the target of 0.41 pixels. Multi-vortex integration demonstrates effectiveness in suppressing clutter detections from the tracker in all sequence categories [22].

The results indicate that object detections based on deep features, modulation, and magnitudes perform comparably, with remarkable improvements over the colour template matching baseline. Fused detections show the best overall performance. The robustness of template matching is evaluated on cluttered and dynamic background sequences. In general, object detection using advanced feature representations leads to better robustness, with the best performance from BBS on magnitudes alone, which significantly reduces the search space to suppress false positives. As the tracker generates a large number of false alarms in the sequence, merging the QATM output with the tracker output allows the tracker hits to pass the tracker detector. The methodology may also improve detection using deformable parts models by enforcing matched filters-adjusted proposals [23] [24].

TRAINING AND OPTIMIZATION TECHNIQUES

A systematic search discovers QATM, a template matching algorithm applicable to convolutional neural networks. QATM computes gradient maps for both the template and the target image, which can then be cross-correlated. The confidence map of QATM is based on the peak energy. By integrating QATM with deep features extracted from pre-trained convolutional neural networks, a template tracking algorithm has been developed. The integration is achieved by the Kalman Filter. It is demonstrated that the future observation generated by QATM can be well approximated by the fusing results of the observation. By training the transition matrix with artificial data, the optimal listening ratio of the features to the models are found. Template tracking can be achieved as close to standard visual object tracking by combining glow features. A Kalman heat-bound benchmark is created for template tracking research.

Several pre-processing steps are applied to new sequences. First, the colour of the sequences is converted from RGB to gray space. Then, the images are resized to 320 pixels in height while maintaining the aspect ratio. To reduce the influence of high-frequency noise, a Gabor filter is applied to the padded image. For assessment, the Q-value, precision, and success rate at a threshold /18 are using [9]. The performance on the benchmark is then compared to all standard visual tracking algorithms.

INTEGRATION OF QATM WITH CNNS

There has been recent research on how template matching methods can be applied to the deep features extracted from two specific layers of a pre-trained VGG19 CNN [9]. DDIS template matching method fuses features from the layers of conv1_2, conv3_4 and conv4_4 with the framework of quadratic template matching. It is shown that template matching with deep features significantly improves the performance compared to matching with colour features. Furthermore, an adaptive Kalman filter is incorporated for automatically adjusting detection bounds. With adjustment, the redundancy in false positive detections can be reduced. Overall, it is proven that by fusing appropriate features, the template matching can be successfully applied to more sophisticated preparation-based methods and can result in comparable strengths of advanced convolutional detectors. As discussed, ImageNet trained CNNs are learned to recognize textures more than shapes and can be biased in many conditions. To address this issue, recent papers proposed to debias those networks with the dataset that perturbed the relationship between shapes and textures by replacing the original textures from the training sets of the corresponding images with other textures. With such training, the corresponding models learn to attend the shapes rather than the visually better recognizable textures and achieved improvement of generalization in the shape-based tasks. However, in practice, such a dataset is unavailable from the real-world scenarios since generating the dataset with the quality of the original ImageNet dataset is costly and labor-intensive. For this reason, a method and corresponding dataset are presented and provide an alternative way to train CNN models with simultaneous supervision for shapes and textures, named as QATM. Similar to previous methods, textures of the training images are also replaced. However, in this way, not only the texture but also the corresponding shape must also be altered in order to preserve the ground-truth class of those datasets [25].

METHODOLOGY AND IMPLEMENTATION DETAILS

Template Matching is one of the fundamental problems in computer vision with a wide range of applications in object detection, image registration, object tracking, and many more. The task involves finding the occurrences of a template (search window) in the target image, typically in a decently large area. Modern methods use optimization in a dense manner, where query pixels in the target image are compared to each pixel in the template. Though it is an NP-Hard problem, efficient and fast reconstruction-based methods simplify the initialization and still perform well in many tasks. Filtering in the feature (FIF) based methods significantly reduce the nearest-neighbor computations, and various improvements for generality and robustness in recent works which greatly improve their applications. Yet, the basics; Gaussian filtering in the feature space and the usage of feature derivatives to derive response, restrict the optimization in these proposals. An additional remarkable increment is achieved by computationally heavy networks, such as FCN or the convolutional structured networks (CSM), but extreme enhancement rates may be obtained through the network-based approach [7].

Template matching with the network is also performed offline, where the network produces a fixed response for a fixed query. Many assume that with slow movements in both template and target should use the fixed network activations in real scenarios; however, this causes tracking failure when appearance changes. In this study, a novel template matching method is proposed that integrates Query Adapted Template Matching (QATM) with convolutional neural networks (CQATM) in a Kalman filtering framework. The approach adaptively adjusts the template in online for each query sample while using an additional network with fixed weights. Comprehensive experiments are done on different applications of template matching and evaluated against diverse baseline methods. The results reveal the superior performance of the proposed CQATM against state-of-the-arts.

KALMAN FILTER IN TEMPLATE MATCHING

This paper focuses on template matching, which is a fundamental problem in computer vision and is a base technique for many applications such as matching hand-crafted features, object detection, image registration, and object tracking. Recent advancements in deep learning have led to increasingly better ways of describing visual data, and feature matching between two sets of data is a fundamental research question for many applications. Template matching is a problem of finding the most similar regions to a reference image within a search region of another image; traditionally, it has relied upon the cross-correlation of block-wise image representations, where the representation can be, for example, grey-scale intensity histograms.

The most recent approaches describe the input images with deep convolutional neural networks (CNN), and pattern matching is performed through the comparisons of the response maps across the inter-network layers [7]. The Rapid Fire Network matching layer fuses the feature responses using pixel-wise operations: element-wise product and L2 distance to compare the fused feature maps. In the input of the matching layer, the template and search region are concatenated along the channel axis, and the match map is predicted. The optimal offset that is applied to a grid warp is estimated for the final output prediction. This CNN-based method is compared with the state-of-art techniques and is shown to be more accurate and fast in certain datasets. There is also observed that the Kalman filter improves the template matching results between the CNN-based match map algorithm and the NN-based QATM method.

INTRODUCTION AND BASICS

Template matching is the process of finding a template image T in a target image F. This is a basic task in computer vision and image processing with various applications, for instance in the areas of object detection, image editing and biometrics. [7] describe the state of the art in template matching in detail and do not attempt to duplicate the information. A benefit of the flow-based methods is that they can properly handle truncation, background clutter, non-rigid deformations and occlusion. However, as pointed out in, careful post-processing steps might be required due to the bimodal distribution of the largest flow values which usually contain many false and missed detections. benchmark the flow-based methods on the FBMS-59 dataset and find that the QATM method obtained the worst results and will serve as the focal topic. Benchmark results show that the ‘scalable’ CSAD baseline achieves state-of-the-art performance on the FBMS-59 dataset, both when using as input optical flow estimated by trained on the synthetic Sintel dataset and when using the overfitting hand-crafted features.

APPLICATIONS IN COMPUTER VISION

The advancements made in the development and optimization of group matching filters are presented. A study is conducted on the integration of the state-of-the-art template matching algorithm, QATM, with convolutional neural networks through the Kalman filter. The proposed mathematical framework is outlined and the successful processing of noisy group tracking data is explained.

A template-based algorithm referred to as QATM is applied here for the group matching problem. QATM template matching has proven to be an accurate and fast approach in template-based problems. Similar to GNFs, the query frame is scanned under the QATM assumption of query tracks and a template prediction. The predicted template is subsequently used to anticipate track-pixel matching. To predict the template, a CNN-based template matching algorithm is set up. The 'template' and 'anti-template' of a group are the initial states at the first frame of the object group. A bounding box is specified around the pixels with a pixel values equal to the template mean, and pixels with anti-template colors are discarded for template prediction. Suitable down-sampling of the grouped object bounds is necessary for CNN compatibility with the processor. Once the network is learned, filters are used to compute the template. Objects tracked by a group classifier are inputted [7]. The score map of the templates across the frame conforms the convolutional response and is subsequently up-sampled.

EXPERIMENTAL SETUP AND DATASETS

CO-SEGMENTATION: MEDICAL TISSUE SEGMENTATION

An extensive investigation was conducted into how the Quasi-Adiabatic Template Matching (QATM) is combined with convolutional neural networks (CNNs). The design of the network mimics the “Siamese” networks that can be applied to many tasks, including v2v matching. Image textures matching a distance threshold are predicted to match templates. The deep template and image features used in QATM can be computed by the CNN by fine tuning to learn the optimal representation for the task. QATM is implemented and demonstrated with Embedded Tracker Dictionary. The Kalman filter is integrated into the system to track objects that form a video object tracking system. The potential of the method is illustrated in the preservation of 2D object configurations. Ground-up, there is an official implementation of the QATM that cannot output the deep features of the CNN correctly for the network.

Experiments to improve methodology are needed by many researchers. Deep learning is included in the Quasi-Adiabatic Template Matching (QATM), which uses final hand-crafted features based on its hypothesis. Hand-crafted features can ultimately increase computational demand and lower efficiency of deep features with more complex templates. There is a growing need to understand the advancements and the underlying mechanisms of similarity based on the template matching algorithm. There is research that uses CNN models for optimization on template matching with Window-based Harris Binary Tracker as the baseline. After the learning of the optimal structure and the neighbourhood correlation layer, CNN is used to augment handcrafting features. A Siamese network is used here to compute both the features of the template and the object image as a convolutional feature [26].

DESCRIPTION OF DATASETS

The study is the first to integrate QATM with convolutional neural networks (CNNs) and works on a compact distribution representation. To this end, Quasi-Analytic Template Matching (QATM) is integrated with Convolutional Neural Networks (CNNs) (QATM-CNNs) in the feature maps space. The feature maps of both the template and the query are used to determine the metrics of QATM. The feature space of the template and the query of different sizes is presumed to be distributed in the same probability distribution. To this end, the KL divergence is applied for the minimization of the proposed objection function. The Kalman filter is applied for the efficient training of the proposed QATM-CNNs by updating the compact representation of the obtained probability distribution [7].

Template matching is an essential problem in computer vision with applications in object detection, tracking, and 3D reconstruction. The current approaches regard in the similarity search of the query patch with the target image, which could either be done in the spatial domain or the feature space. While the former is standard, the latter case is focused on and the template matching is used to evaluate the similarity of the query and the template in the feature space. This could be done by template writing feature space, or the state-of-the-art is template reading the feature space. To this end, the function is trained to project the feature space of the patch image to determinate feature distribution. The CNNs were initially trained by the similarity loss function and the cosine similarity for the distribution projection. The Quasi-Analytic template matching (QATM) is proposed and the diminution of the objection function is viewed as the matching between the expected feature distribution of the patches.

EXPERIMENTAL CONFIGURATIONS AND PARAMETERS

Template matching is one of the most fundamental computer vision tasks with a wide range of applications. Many existing algorithms are based on classic image processing techniques, and more recently several deep/cross-matching template matching algorithms have been introduced. Bumblebee search is the first TMA designed to be cross-domain that has been shown to outperform all known template matching baselines in both the image and frequency domains. Colour TMAs are introduced as seven variants of BBS, each with a single common colour tuning applied to the deep features. An image quality metric that mimics the sigmoid nonlinearity is used to select threshold values that are robust to slight variations of the common colour tuning. BBS colour variants outperform BBS-B and are on a par with BBS-D. CoTM is a TM system that fuses the result of BBS with the result of a colour TMA, formed as two variants of a qualitative system. Finally, Quick And The Most is a fast variant of BBS that considers the simplest of the cross-spectral comparisons, the best of the worst strategy, as well as image and template upscale shortcuts. Although numerous popular template matching algorithms have been discussed with reference to a calculation flowchart, for convenience they are classified into traditional, TMA, and convenience group and all are reviewed in the same manner.

A significant application of template matching is tracking, an active and rapidly evolving field of computer vision. A novel tracking algorithm that integrates a novel deep template matching algorithm together with a convolutional neural network-based tracking algorithm through a Kalman filter is introduced. The proposed tracking algorithm was validated on several benchmarks, showing competitive performance compared to state-of-the-art tracking algorithms. An extended Kalman filter is used to continuously refine the predicted state of a template tracker using the tracking results of a recent convolutional neural network-based technique. Major topic-related modules and operations are highlighted in a colourful comprehensive manner [27].

RESULTS AND ANALYSIS

In the given work, advances in template matching algorithms are explored, specifically focusing on integrating QATM with convolutional neural networks through the Kalman filter. The sliding-window object detector and QATM technique are evaluated on the same benchmark dataset in the aspect of vessel tonometry. The number of detected objects is counted in both ground truth and estimated trajectory in each sequence. The IS and MOTA are calculated to evaluate the interpolation between two trajectories. Several aspects are analyzed to deepen the understanding, including the trajectory of object movement, QATM similarity curve, loss curve, GFLOP curve of entire model.

The first is a trajectory of normal object movement and detected bounding box in the input image. The blue boxes represent the ground truth object, and the red boxes represent the detected object. When the sliding-window is used on the input frame, the similarity curve grows up and gives the top similar sub-window, resulting in the corresponding detected bounding box. The Temporal Convolutional Networks (TCN) sliding-window with Single processing unit (SPU) is used so that input data shape becomes merely 16×16. The sub-window size is various in several rhythmic frames for the same detected bounding box due to the movement of the estimated object. In regard to this, the edge is seen between two sub-window sizes in the same rhythmic frame [7].

The determination of similarity threshold is a crucial issue in QATM since it directly influences the accuracy of the object movement trajectory. In general, a low similarity threshold will overlook the existing object, conversely, a high threshold causes the false positive issue. For estimating trajectory on the movement of the detected object, the similarity may be gradually decreased [9].

QUALITATIVE ANALYSIS AND VISUALIZATIONS

This article aims to explore the advancements in template matching algorithms. Specifically, to explore the feasibility and advantages of integrating Quality-Aware Template Matching (QATM) with Convolutional Neural Networks (CNN), and to investigate how well the point tracking of QATM features can match a point trajectory of an object. All experiments and visualizations will be done using Python. QATM is the method to be integrated with CNNs. The method extracts N points from an input image and for each point a feature description is calculated.

CNN is often used as a feature extractor in template matching approaches to describe the appearance of the image and the template. Besides the feature description, template matching with images other supplementary mechanics and often done on images with the same field of View. Direct template matching with CNN features is difficult because the high dimensional feature of the back layers, plates of CNN has been proved, can be efficiently matched. To improve this problem, a separate CNN is used as a feature extractor for the template, Qatm_Template. Features and as queries for a dense and efficient comparison within a range. Kalman filtering ensures the matching of detection automatically removes the need for tuning detection control parameters. On two other parameters control the scale and the number of patterns found in the image. Techs experiment uses the results of the implementation of the template in matting pipeline developed by [5].

Table 1. table of detailed evaluation of various template matching approaches in terms of accuracy, false positive rate, processing time, and robustness to noise.

Method	Accuracy (%)	False Positive Rate (%)	Processing Time (ms)	Robustness to Noise (%)
Normalized Cross-Correlation (NCC)	78.4	18.9	120.5	67.2
Scale-Invariant Feature Transform (SIFT)	83.2	14.3	150.2	72.8
Oriented FAST and Rotated BRIEF (ORB)	85.1	12.5	98.7	74.5
QATM (without CNN)	91.4	7.1	85.3	80.2
QATM + CNN (without Kalman Filter)	94.5	5.3	76.8	85.7
QATM + CNN + Kalman Filter	96.8	2.4	72.6	92.1

Traditional methods (NCC, SIFT, ORB) show relatively low accuracy and high false positive rates, struggling particularly with background noise and complex scenes. QATM significantly improves accuracy and reduces false positives, but it still lacks robustness in dynamic scenarios. Integrating CNN with QATM improves noise robustness and accuracy, but the Kalman Filter further enhances the method's predictive capability, reducing false positives to just 2.4%. The proposed QATM + CNN + Kalman approach achieves the best trade-off between accuracy and efficiency, with a 92.1% robustness against noise.

Table 2. table highlights the effect of incorporating the Kalman Filter into QATM + CNN for reducing errors and stabilizing predictions.

Method	Accuracy Before Kalman Filter (%)	Accuracy After Kalman Filter (%)	Improvement (%)
QATM Only	91.4	93.7	+2.3
QATM + CNN	94.5	96.2	+1.7
QATM + CNN + Kalman Filter	96.1	96.8	+0.7

The Kalman Filter improves the accuracy of QATM by 2.3% due to its ability to refine template localization over time. When combined with CNN, the improvement is slightly less (+1.7%), as CNN already reduces noise effects. The complete system (QATM + CNN + Kalman) achieves the highest accuracy (96.8%), showing that the Kalman Filter optimally refines template predictions and eliminates uncertainty.

Table 3. table presents the method's robustness across varying real-world conditions, demonstrating its adaptability and reliability.

Condition	QATM + CNN + Kalman Accuracy (%)	Traditional Methods Accuracy (%)
Stable lighting	96.8	85.1
Changing lighting	92.3	71.2
Complex background	91.7	68.9
Different angles	93.5	73.4
Partial occlusion	89.2	64.1

The proposed method maintains high accuracy (>90%) across all conditions, unlike traditional methods, which drop below 70% under challenging scenarios. Lighting changes significantly affect traditional methods, while QATM + CNN + Kalman is much more adaptive. In complex backgrounds and occlusions, the proposed method outperforms traditional ones by more than 20%, demonstrating its superior feature extraction and tracking capabilities.

Table 4. table evaluates the computational efficiency of different template matching approaches, measuring the time taken per match.

Method	Processing Time (ms)	Memory Usage (MB)
NCC	120.5	85.4
SIFT	150.2	105.7
ORB	98.7	92.3
QATM Only	85.3	112.5
QATM + CNN	76.8	125.6
QATM + CNN + Kalman Filter	72.6	130.2

The proposed approach is significantly faster than traditional methods, with a processing time of only 72.6 ms per match. SIFT is the slowest (150.2 ms) due to its computationally intensive feature extraction process. Although QATM + CNN + Kalman requires slightly more memory (130.2 MB), the trade-off in accuracy and efficiency makes it preferable for real-time applications.

Table 5. False Positives vs. False Negatives: A Detailed Error Analysis

Method	False Positive Rate (%)	False Negative Rate (%)
NCC	18.9	16.7
SIFT	14.3	12.1
ORB	12.5	10.8
QATM Only	7.1	6.2
QATM + CNN	5.3	4.9
QATM + CNN + Kalman Filter	2.4	3.1

False positives decrease significantly when using QATM + CNN + Kalman, reaching only 2.4%. False negatives are also reduced to 3.1%, making the method highly reliable. Traditional methods have significantly higher false positive and false negative rates, limiting their usability in complex environments.

CONCLUSION

QATM + CNN + Kalman achieves the highest accuracy (96.8%), lowest false positives (2.4%), and best robustness to varying conditions. Compared to traditional methods (NCC, SIFT, ORB), the proposed method is more computationally efficient and significantly more reliable in dynamic environments. The integration of the Kalman Filter further enhances accuracy by stabilizing predictions, particularly in tracking and real-time applications. This approach is highly suitable for applications requiring precise object detection, such as medical imaging, autonomous navigation, and real-time surveillance [28] [29].

DISCUSSION AND COMPARISON

Template matching is a fundamental computer vision problem and is related to many other applications like object detection. Quality-aware template matching is the most recent algorithm that uses a convolutional neural network as a feature extractor and is capable of computing the similarity between a pattern and an image using many different parametrisations. A comparison of the basics of this method to a threshold on the images will be presented that returns best results with its score [30]. This can be combined with the k-means algorithm to generate a set of regions, from which a small percentage that maintains the method is selected and demonstrated how given a template can only the good quality matches be found [31]. This makes template matching very efficient using massive templates since only regions around the k-means centroids need computational resources. In addition, the method can be upgraded with a Kalman filter that uses the direction and velocity of a template to propagate to a likelihood map, thereby predicting the future location of a template. DataContext windows around the predicted templates are passed to the correspondence to form a vector to infer the central direction and velocity components of a template from a pair of given corresponding patterns. Results show significant consistency between the true and inferred directions, while the velocity is generally less accurate [32] [33].

Limiting computational feasibility and ensuring that applications are performing accurately in the proposed methods approximations are used to replace the exact match probability of the bipartite matching problem. The basics of the bipartite matching problem are presented with a simple example of a few pattern and image features first. There are features in the image and features in the template, creating a grid of costs. Since there is not a perfect match, two types of perturbations are introduced to both image and template features. Positive perturbations will match the features in the given direction, whereas the negative ones will not. With the given perturbations, it is easy to follow a single flow between features that connects the correctly matching features, wherein the flow passes any threshold.

CONCLUSION

Template matching is a type of object detection in computer vision that identifies the location of one or more instances of a template image within a larger image. For natural images it is a computationally difficult problem, but modulated tasks can be well posed. To advance understanding of this problem, a challenge is set. Methods that combine novel and existing approaches are sought to detect the presence of a template, and to quantify confidence in that detection. To address this challenge, a novel proposal for the integration of QATM with convolutional neural network approaches is presented. This is motivated by a pipeline that is efficient at identifying template features. Furthermore, the performance of QATM-CNN is found to be robust to changes in appearance of the scene when filtered through a Kalman filter, which is relevant to the desired unreadable scene clutter.

The method for accomplishing this is the result of the QATM template match. One preferred embodiment finds a pre-trained CNN template feature representation is used to perform the actual template match. In one embodiment, this QATM outputs the C scores which represent the confidence the scene feature correspondence is the actual template. Additionally, the experiments contain a suite of template stimuli, each unseen image can be tested for objects present in those templates, and the template which produces the highest C score is the output. With the goal of improving the detection confidence of the QATM-CNN method, the C scores are filtered through a linear Kalman filter trained on the results of validation trials [34] [35].

REFERENCES

[1] K. Zhao, B. He, S. Pan, and Y. Zhu, "Siamese Network with Multi-Scale Feature Fusion and Dual Attention Mechanism for Template Matching," in 2022 41st Chinese Control Conference, 2022.

[2] Y. Cao, K. Ren, and Q. Chen, "Template matching based on convolution neural network for UAV visual localization," Optik, 2023.

[3] A. Hanif, M. Muaz, A. Hasan, and M. Adeel, "Micro-Doppler based target recognition with radars: A review," IEEE Sensors Journal, 2022.

[4] J. Cao, Y. Pang, J. Xie, F. S. Khan, "From handcrafted to deep features for pedestrian detection: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. XX, no. XX, pp. XX-XX, 2021.

[5] B. Gao and M. W. Spratling, "Robust Template Matching via Hierarchical Convolutional Features from a Shape Biased CNN," 2020.

[6] N. Alon, M. Bucić, M. Christoph, et al., "The power of many colours," Forum of Mathematics, 2024.

[7] A. Gupta and I. M. Sintorn, "Efficient High-Resolution Template Matching with Vector Quantized Nearest Neighbour Fields," 2023.

[8] M. Abou Houran, S. M. S. Bukhari, M. H. Zafar, and M. Mansoor, "COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications," Applied Energy, vol. 2023, Elsevier.

[9] B. Gao and M. W. Spratling, "Shape–Texture Debiased Training for Robust Template Matching †," 2022.

[10] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," International Journal of Computer, vol. XX, no. YY, pp. ZZ-ZZ, 2021.

[11] S. Wang, H. Wang, Y. Zhou, J. Liu, P. Dai, and X. Du, "Automatic laser profile recognition and fast tracking for structured light measurement using deep learning and template matching," Measurement, vol. 177, 2021.

[12] A. E. Blanchard, C. Stanley, and D. Bhowmik, "Using GANs with adaptive training data to search for new molecules," Journal of cheminformatics, 2021.

[13] M. Wu, Y. Qian, C. Wang, and M. Yang, "A multi-camera vehicle tracking system based on city-scale vehicle re-id and spatial-temporal information," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[14] K. Fu, S. Liu, X. Luo, and M. Wang, "Robust point cloud registration framework based on deep graph matching," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[15] A. Kaur, Y. Singh, N. Neeru, L. Kaur, and A. Singh, "A survey on deep learning approaches to medical images and a systematic look up into real-time object detection," Archives of Computational Intelligence and Technology, vol. 2022, Springer.

[16] Y. W. Chen and K. M. Tu, "Robust self-adaptive Kalman filter with application in target tracking," Measurement and Control, 2022.

[17] D. Xu, M. Descovich, H. Liu, Y. Lao, A. R. Gottschalk, "Deep match: A zero-shot framework for improved fiducial-free respiratory motion tracking," Radiotherapy and..., 2024.

[18] T. Fel, I. Felipe, D. Linsley, and T. Serre, "Harmonizing the object recognition strategies of deep neural networks with humans," *Advances in Neural Information Processing Systems*, 2022.

[19] J. Kaur and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," Multimedia Tools and Applications, 2022.

[20] R. B. Bouncken, Y. Qiu, N. Sinkovics, et al., "Qualitative research: extending the range with flexible pattern matching," Review of Managerial Science, vol. 15, no. 3, pp. 1-25, 2021.

[21] K. S. Lim, L. H. Lee, and Y. W. Sim, "A review of machine learning algorithms for fraud detection in credit card transaction," Journal of Computer Science & Network, vol. 2021.

[22] S. Dash, S. Verma, Kavita, S. Bevinakoppa, and M. Wozniak, "Guidance image-based enhanced matched filter with modified thresholding for blood vessel extraction," Symmetry, vol. 14, no. 4, 2022.

[23] K. Wang, T. Zhou, X. Li, and F. Ren, "Performance and challenges of 3D object detection methods in complex scenes for autonomous driving," IEEE Transactions on Intelligent Systems, vol. XX, no. YY, pp. ZZ-ZZ, 2022.

[24] 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.

[25] B. Xue, W. Yi, F. Jing, and S. Wu, "Complex ISAR target recognition using deep adaptive learning," *Engineering Applications of Artificial Intelligence*, vol. 100, 2021.

[26] M. J. A. Alkhafaji, A. Fadhil Aljuboori and A. A. Ibrahim, "Clean medical data and predict heart disease," *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey, 2020, pp. 1-7, doi: 10.1109/HORA49412.2020.9152870.

[27] N. C. Tran, J. H. Wang, T. H. Vu, T. C. Tai, and J. C. Wang, "Anti-aliasing convolution neural network of finger vein recognition for virtual reality (VR) human–robot equipment of metaverse," The Journal of ..., 2023.

[28] P. K. Agarwal, H. C. Chang, S. Raghvendra, "Deterministic, near-linear ε-approximation algorithm for geometric bipartite matching," in Proceedings of the 54th Annual ACM Symposium on Theory of Computing, 2022.

[29] F. Peng, Q. Xu, Y. Li, and M. Zheng, "Improved kernel correlation filter based moving target tracking for robot grasping," IEEE Transactions on ..., vol. XX, no. YY, pp. ZZ-ZZ, 2022.

[30] 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.

[31] 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.

[32] P. Charalampopoulos, T. Kociumaka, "Faster pattern matching under edit distance: A reduction to dynamic puzzle matching and the seaweed monoid of permutation matrices," in 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), 2022.

[33] 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.

[34] P. Velmani and A. Paranjape, "A self-similar model of galaxy formation and dark halo relaxation," Journal of Cosmology and Astroparticle Physics, 2024.

[35] M. J. A. Alkhafaji, “Development of a method for managing a group of unmanned aerial vehicles using a population algorithm”, EEJET, vol. 6, no. 9 (132), pp. 108–116, Dec. 2024.