
[ICCV 2025] Decision for Submission #2220

보낸 사람 OpenReview <noreply@openreview.net>

날짜 목 2025-06-26 오전 2:50

Dear Authors of Submission #2220,

Thank you for your submission to ICCV 2025. The review process has now concluded. Below you will find the meta-review and final reviews for your submission, which will be available on OpenReview shortly.

Congratulations, your submission #2220, titled "Few-Shot Pattern Detection via Template Matching and Regression" has been accepted to ICCV 2025.

You will receive additional information for submitting the camera-ready version shortly. Please note that acceptance is contingent on passing a plagiarism check. Papers that do not comply with plagiarism and dual-submission rules will be rejected.

This year, we received 11,239 valid submissions that underwent the review process. The program committee recommended 2,698 papers for acceptance, resulting in an acceptance rate of 24%. All papers were initially evaluated by at least three independent reviewers. Following the author rebuttal and reviewer discussions, moderated by an Area Chair (AC), a triplet of three ACs reviewed each paper holistically, considering the reviews, author rebuttal, and reviewer discussions. For challenging cases, additional ACs and/or Program Chairs (PCs) were consulted.

We would like to thank the reviewers and Area Chairs for their contributions to the review process.

Best Regards,
ICCV Program Chairs

Metareview

Paper receives pre-rebuttal ratings of BA, BR, BA. After the rebuttal, the BR turns around and recommends BA, and one BA upgrades to WA, so I am accepting the paper with post-rebuttal BA, BA, WA ratings. Please address the remaining sentiments from the original BR reviewer in the final version.

Final Recommendation

Accept

Reviewer oRyb**Paper Summary**

This paper introduces a few-shot pattern detection method called Template Matching and Regression (TMR). It focuses on detecting repeated patterns, including non-object patterns, using a simple architecture that avoids collapsing spatial information, which is common in existing few-shot detection methods. The authors also present a new dataset, RPINE, designed to include a wider range of repeated patterns beyond objects. The proposed method outperforms existing approaches on RPINE, FSCD-147, and FSCD-LVIS, especially in cross-dataset generalization.

Paper Strengths

- The paper addresses a clear limitation in current few-shot detection work, which mostly focuses on objects with clear semantics and boundaries.
- The TMR approach is simple, using classic template matching and regression.
- The new RPINE dataset is a meaningful contribution. It includes non-object patterns and supports evaluation across a wider variety of visual repetition.
- Experimental results are strong, especially in cross-dataset settings, suggesting that the method generalizes better than those that rely on object-level features.

Major Weaknesses

T is extracted with ROIAlign. I assume this is to get a fixed height and width for T. But this transformation is not applied to F, so the correlation between T & F will miss cases where ROIAlign drastically changes the spatial arrangement. In general, it seems the system wasn't designed to be invariant to scale/rotation, which seems to be an important requirement for most use cases. There is only so much the later conv layers can do to overcome this. This is also evident in Fig 5, where different scale matches are not detected.

While the model head is light, it relies on a large, frozen SAM-ViT backbone. This dependence contradicts some of the claims about simplicity and efficiency (e.g. "minimalistic architecture, which consists of a few learnable layers").

Feature upsampling from 64×64 to 128×128 doesn't add any new info. Wouldn't it be better to work in 64×64 and allow the predicted/GT bounding boxes to be sub-"pixel"?

Minor Weaknesses

- Section 3 is very short and can be a subsection
- In Fig. 5, mention what the blue BB indicate

Preliminary Recommendation

4: Borderline Accept

Preliminary Justification

The paper takes on an important challenge in few-shot detection by focusing on non-object patterns, and the results show meaningful improvements, especially in generalization. The method is straightforward, and the new dataset adds value. My main concern is the lack of invariance to basic visual transformations.

Confidence Level

4 - High Confidence - The reviewer has strong expertise in the area. They are highly familiar with the relevant literature and can critically evaluate the paper.

Final Recommendation

5: Weak Accept

Final Justification

The authors provided satisfactory responses to my concerns.

Reviewer rfZa

Paper Summary

The paper presents a pattern detection model based on a single-stage object detection framework. By computing convolutional correlations between template exemplars and query features, it predicts bounding boxes and objectness scores. These detections are subsequently refined using SAM.

Paper Strengths

1. The paper is relatively easy to follow, although some technical details remain unclear.
2. Extensive experiments are conducted, showing competitive results.

Major Weaknesses

1. Since the convolution correlation and regressing Δx and Δy are standard operations in Faster R-CNN, the template detection without SAM is more like training an expert model for template detection. The novelty in architecture is limited, despite the results being competitive.
2. From Table 4, the performance TMR without SAM is also competitive. Does not mean that the expert model can work better on this task and the general model is not needed? I would like to see SAM-free results on other datasets, such as FSCD-LVIS.
3. Unclear details about the framework and the scale of parameters:
 - What is the backbone used in Fig.4? Is it SAM's encoder? If the backbone is SAM's encoder, as claimed in the paper it is frozen during training. In this way, only the box regressor is trainable. Then what is the exact structure of the regressor? How large is it?
 - If the backbone in Fig.4 is not SAM's encoder, what is the backbone exactly?
 - Ultimately, how large is the expert model shown in Fig.4?
3. For Eq.6, is there a margin between positives and negatives? Following Eq.6, will the model become confused if the point falls on the boundary of the GT region?
4. For the new dataset, the statistics shown in Table 1 are ambiguous. How to evaluate pattern tiling and repetition, and object bias? It is not supported by numbers. In addition, in section 5, it lists some perspectives, but they are not supported by statistics in Table 1, while Table 1 presents statistics in other perspectives which are unexplained in the text.

Minor Weaknesses

1. In Eq1, does it need to subtract $t_w/2$ and $t_h/2$?
2. Expression errors in In p5. L 314: weight and height lower bound --> the width and height of lower bound of

Preliminary Recommendation

3: Borderline Reject

Preliminary Justification

I expect a clear explanation about the framework and the model size.

Confidence Level

3 - Moderate Confidence: The reviewer is reasonably knowledgeable about the topic. They understand the paper's methodology and results.

Final Recommendation

4: Borderline Accept

Final Justification

The reviewers have addressed most of my concerns. Even though the novelty is limited, considering its good performance and contribution of a new dataset, I would like to raise the score to BA. Despite this, I am still curious as to why previous designs of existing SOTAs cannot outperform the conventional regression and matching approach, as tried in this work. I would like the authors to provide more deep analysis on this point, especially when the method is simple and straightforward. In addition, Figure 2 is not entirely consistent with the framework described in the text. The text mentions a decoder, while Figure 2 does not.

Reviewer USvT

Paper Summary

This paper addresses the problem of few-shot pattern detection, a generalization of few-shot detection/counting tasks to include non-object visual patterns (e.g., textures, abstract elements). The authors propose a lightweight and intuitive method called Template Matching and Regression (TMR), which leverages 2D template matching between exemplar and input image features and learns to regress boxes using a minimal architecture (3×3 convs + projections). A new dataset, RPINE, is also introduced to benchmark this broader pattern detection task. TMR achieves state-of-the-art results on RPINE, FSCD-147, and FSCD-LVIS, with notably strong generalization to unseen datasets.

Paper Strengths

Task reformulation: This paper clearly motivates and formalizes a more general version of few-shot detection, moving beyond object-centric detection to arbitrary pattern matching — an important and underexplored problem.

Simplicity and efficiency: TMR uses a straightforward architecture that forgoes complex modules (e.g., cross-attention) yet performs competitively with lower FLOPs and parameter count.

Strong generalization: TMR shows superior performance in cross-dataset evaluation (e.g., trained on RPINE → evaluated on FSCD-147), highlighting its robustness across domains.

New benchmark (RPINE): The introduction of RPINE fills a gap in existing datasets, covering diverse, weakly-structured, and non-object visual patterns with multi-pattern annotation per image.

Comprehensive experiments: Results are provided for both detection and counting tasks, on multiple datasets, with strong ablation and efficiency comparisons.

Major Weaknesses

Limited architectural novelty: While TMR is effective, its underlying method — template matching followed by local regression — is conceptually simple and builds closely on classical ideas.

Lack of visualizations for failure cases: The paper focuses on performance gains, but misses analysis of when and why TMR fails, especially in cluttered or highly textured regions.

Limited evaluation on real-world applications: Despite RPINE's novelty, the paper lacks downstream application studies (e.g., industrial quality inspection, scientific pattern analysis).

SAM decoder criticism feels shallow: The claim that the SAM decoder harms performance is valid, but not thoroughly studied — additional failure cases and comparisons would strengthen this argument.

Minor Weaknesses

Include qualitative visualizations of failure cases and misaligned predictions (especially non-object patterns).

Provide more insight into the limitations of RPINE (e.g., annotation subjectivity, bounding box ambiguity).

Discuss possible extensions to multi-scale or dense correspondence tasks — how could TMR adapt?

Consider adding metrics for pattern diversity or proposing new metrics beyond AP/MAE that better reflect the pattern-matching nature of the task.

Preliminary Recommendation

4: Borderline Accept

Preliminary Justification

This paper is a timely contribution to few-shot visual learning. It generalizes the detection task in a principled way and introduces a new dataset to benchmark a more realistic problem. While the method itself is relatively simple, its performance, efficiency, and generalization capacity are impressive. The paper is well-written and experimentally thorough, and I believe it will be of interest to the ICCV community.

Confidence Level

3 - Moderate Confidence: The reviewer is reasonably knowledgeable about the topic. They understand the paper's methodology and results.

Final Recommendation

4: Borderline Accept

Final Justification

I appreciate the efforts by authors and would like to keep my positive score as they have addressed concern properly.
