*Microscopy and Microanalysis*, **29** (Suppl 1), 2023, 931
https://doi.org/10.1093/micmic/ozad067.462

Microscopy AND Microanalysis

**Meeting-report**

# Exploring the Limits of 2D Template Matching for Detecting Targets in Cellular Cryo-EM Images

**Kexin Zhang[1,2], Bronwyn Lucas[3], and Nikolaus Grigorieff[1,2,*]**

[1]*The University of Massachusetts Chan Medical School, RNA Therapeutics Institute, Worcester, MA, USA*
[2]*Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA, USA*
[3]*Department of Molecular & Cell Biology, The University of California, Berkeley, Berkeley, CA, USA*

*Corresponding author: niko@grigorieff.org



Accurately placing macromolecular assemblies in the cellular context is an important step in understanding their mechanistic role inside the cell. Template matching is a classic object detection method that finds patterns in an image that match a template by calculating a cross-correlation (CC) at each x and y location in the image. The CC output will be the highest at locations that contain target patterns. Template matching has been successfully applied to locate particles *in situ* by matching tomograms with a 3D reconstruction of the target (3DTM) [1]. However, the inherent low resolution of the tomogram and 3D template can cause a high false positive rate in 3DTM unless the further classification is performed [2]. 3DTM relies largely on low-resolution features, the molecular shapes, making it challenging to distinguish targets with compositional or conformational heterogeneity, detect particles in close proximity to one another, or identify particles in densely packed environments, such as nuclei.

Recently, a 2D template matching (2DTM) approach was developed by our lab to locate targets such as ribosomes and viruses in cellular cryo-EM images with high positional and orientational accuracy [2, 3], leading to a complete picture of local environments and molecular structure of a cell. In previous work, we have shown that 2DTM could distinguish related molecular populations such as mature 60S and late nuclear (LN) 60S using the subcellular localization as prior information or distinguish mixed populations of nuclear pre-60S when no prior is available [4].

Despite its success in detecting and discriminating related ribosome intermediates in cells, the current 2DTM approach suffers from several limitations. For example, 2DTM has a high false negative rate, making it difficult to detect rare targets, as it uses a high threshold to discriminate between true positives and false positives. Moreover, multiple factors, including factors unrelated to the structure of the targets, may affect the 2DTM SNRs, complicating their interpretation. Finally, the 2DTM SNRs are proportional to the molecular weight of the target, limiting current detection to 300-400kDa in samples with 150-200nm ice thickness [3, 5]. Targets with smaller molecular weights or thicker ice are more challenging to detect. In this work, we have modeled the signal and noise distribution of 2DTM SNRs using the extreme value theorem and explored the physical limits of 2DTM to better detect targets in cells.

The 2DTM SNR at an x and y location is the maximum CC calculated at this location between the image and template projections in different orientations. The fact that the SNR is defined as a maximum enabled us to model its distribution using the extreme value theorem [6-8]. We have modeled the 2DTM SNRs of 10 micrographs of yeast cells, searched with mature 60S and LN 60S, respectively [4]. Our results showed that the x and y locations in a cryo-EM image could be modeled using a mixture of two populations, locations with signal and without signal. Moreover, the weights of the populations provide an unbiased estimation of target molecule concentration in the cell compared to the number of peaks that are currently used in 2DTM, especially for targets with weaker 2DTM signals (e.g., targets that are more dynamic or with smaller molecular weights).

## References

1. J Mahamid *et al.*, *Current Opinion in Structural Biology* (2018), p. 111. doi:10.1016/j.sbi.2018.08.009.
2. B Lucas, *et al.*, *Elife* (2021), p. e68946. doi:10.7554/eLife.68946.
3. JP Rickgauer, *et al.*, *Elife* (2017), p. e25648. doi:10.7554/eLife.25648.
4. B Lucas, *et al.*, *Elife* (2022), p. e79272. doi:10.7554/eLife.79272.
5. JP Rickgauer, *et al.*, *bioRxiv* (2020), p. 2020.04.22.053868. doi:10.1101/2020.04.22.053868.
6. M Fréchet, *Ann. Soc. Math. Polon* (1927), p. 93.
7. RA Fisher and LHC Tippett, *Proceedings of the Cambridge Philosophical Society* (1928), p. 180. doi:10.1017/S0305004100015681.
8. M Falk and F Marohn, *The Annals of Probability* (1993), p. 1310. doi:10.1214/aop/1176989120.