Image-derived generative modeling of pseudo-macromolecular structures – towards the statistical assessment of Electron CryoTomography template matching

Kai Wen Wang[1]
kaiwenw1@andrew.cmu.edu
Xiangrui Zeng[1]
xiangruz@andrew.cmu.edu
Xiaodan Liang[1]
xiaodan1@andrew.cmu.edu
Zhiguang Huo[2]
zhuo@ufl.edu
Eric P. Xing[1]
epxing@cs.cmu.edu
Min Xu (Corresponding Author)[1]
mxu1@cs.cmu.edu

[1] Carnegie Mellon University
Pittsburgh, PA, USA
[2] University of Florida
Gainesville, FL, USA

Abstract

Cellular Electron CryoTomography (CECT) is a 3D imaging technique that captures information about the structure and spatial organization of macromolecular complexes within single cells, in near-native state and at sub-molecular resolution. Although template matching is often used to locate macromolecules in a CECT image, it is insufficient as it only measures the relative structural similarity. Therefore, it is preferable to assess the statistical credibility of the decision through hypothesis testing, requiring many templates derived from a diverse population of macromolecular structures. Due to the very limited number of known structures, we need a generative model to efficiently and reliably sample pseudo-structures from the complex distribution of macromolecular structures. To address this challenge, we propose a novel image-derived approach for performing hypothesis testing for template matching by constructing generative models using the generative adversarial network (GAN). We conducted hypothesis testing experiments for template matching on both simulated and experimental subtomograms, allowing us to conclude the identity of subtomograms with high statistical credibility and significantly reducing false positives. Our general approach can be extended to assess other template matching tasks, such as face, iris, and fingerprint scans, where statistical tests are also very important.

1 Introduction

The cell is the basic structural and functional unit of all living organisms. Biochemical processes of the living cell are often catalyzed by tiny cellular machines called macromolecular complexes. To fully understand these cellular processes, it is extremely helpful to systematically extract the structure and spatial organization of macromolecular complexes in single cells. Cellular Electron CryoTomography (CECT) [2] is a powerful 3D imaging tool that enables the study of sub-cellular structures at near-native state and in sub-molecular resolution. However, the quality of the reconstructed CECT images (a.k.a. tomograms) suffer from many current imaging limitations, such as low signal-to-noise ratio (SNR), missing wedge and limited number of angular samples, to a point where interpretation by visual inspection is impractical. As such, locating instances of macromolecular complexes inside tomograms has remained an extremely challenging computer vision problem [4]. A popular method for this task has been template matching, which we describe now.

Let $\mathcal{S}_{known}$ denote the set of known macromolecular complexes, let P denote a cubic subvolume of a tomogram containing a single macromolecule (a.k.a. subtomogram), and let $T(C)$ denote the template derived from a macromolecule C. Given P and a known complex $KC \in \mathcal{S}_{known}$, template matching is a method for deciding whether P contains an identical macromolecule as KC by using the Pearson cross-correlation score $c(P,T(KC))$ [6]. To calculate the score, $T(KC)$ is first rigidly aligned against P using a fast alignment method[6, 21] and then the score is computed with missing wedge compensation [5].

In most cases, template matching makes the decision based on a single chosen threshold for the correlation score or the complex most similar to P is declared a match by default from having the highest score from a batch of known macromolecules. However, as cross-correlation is only a relative measure of similarity, this approach is neither rigorous nor statistically meaningful. Instead, hypothesis testing is preferred for quantitatively assessing the statistical credibility of the decision. Performing an accurate hypothesis test is extremely challenging since the number of macromolecules with known structures is very limited. In addition, the structural distribution of macromolecules is highly complex. Across space and time, macromolecules typically adopt different conformations as part of their function and dynamically interact with other macromolecules [22]. Across different species and cell types, the majority of macromolecules are still unknown [9].

In this paper, we propose a statistically rigorous treatment of template matching using a novel, image-derived, Monte Carlo approach for performing hypothesis testing and calculating empirical p-values[10]. The Monte Carlo samples are pseudo-macromolecular complexes (denoted henceforth as C_0) sampled from the structure distribution $f_{structure}$ of macromolecules. We used a 3D-WGAN to learn $f_{structure}$ from a collection of known complexes (see Section 2.1). Our procedure is as follows:

(Step 1) Train a 3D-WGAN to learn the structural distribution $f_{structure}$.

(Step 2) Determine the macromolecule of interest $C_{Interest} := \arg\max_{KC \in \mathcal{S}_{known}} c(P,T(KC))$ to be the known macromolecule with the highest alignment score.

(Step 3) Using pseudo-complexes $C_0 \sim f_{structure}$ generated by the 3D-WGAN and sampled away from $C_{Interest}$ (see Section 2.2), perform a hypothesis test for whether P contains a macromolecule identical to $C_{Interest}$ and calculate the empirical p-value under the null hypothesis H_0 (see Section 2.3).

To formally establish the hypothesis testing procedure, we specify the null hypothesis H_0 and the alternative hypothesis H_A as the following:

$$H_0 : P \text{ does not contain a macromolecule identical to } C_{Interest} \quad (1)$$

$$H_A : P \text{ contains a macromolecule identical to } C_{Interest} \quad (2)$$

The statistical credibility is assessed by the p-value, which is the probability of obtaining $c(P,T(C_0))$ at least as extreme as $c(P,T(C_{Interest}))$, given that H_0 is true. The lower the p-value, the stronger the evidence against H_0, giving more statistical credibility that H_A is true. To calculate an empirical p-value, we randomly sample pseudo-macromolecular complexes $C_0 \sim f_{structure}$ to derive a Monte Carlo empirical distribution of the test statistics $c(P,T(C_0))$ under the null.

Throughout this paper, let $\mathcal{S}_{pseudo}$ denote the set of pseudo-complexes (i.e. macromolecules generated from the 3D-WGAN). We emphasize that $\mathcal{S}_{pseudo}$ and $\mathcal{S}_{known}$ have important differences:

- The identity of $KC \in \mathcal{S}_{known}$ is known and $T(KC)$ is used to identify the macromolecule in a subtomogram through the method of template matching.
- The identity of $C_0 \in \mathcal{S}_{pseudo}$ is unknown and $T(C_0)$ is only used for hypothesis testing.

As a recent advancement in unsupervised deep learning, the generative adversarial network (GAN) learns the distribution of training images on the image manifold and generates highly realistic images [7]. The training process of the GAN is akin to a minimax game between two neural network adversaries, the generator and the discriminator. The generator seeks to improve its output images by minimizing the discriminator's classification accuracy while the discriminator seeks to maximize its accuracy. After training, the generator can produce diverse and realistic images from the distribution of the original training images. To our knowledge, no method exists for constructing such generative models for template search for CECT data.

Our contributions are summarized as follows:

(i) A novel, Monte Carlo approach for statistical assessing template matching through hypothesis testing to calculate empirical p-values.

(ii) An approach for generating the density maps of pseudo-macromolecular complexes, using the 3D Deep Convolutional Wasserstein GAN (3D-WGAN). We showed that our model is able to capture the shape manifold of macromolecules, and sample realistic and diverse pseudo-complexes from the structural distribution of macromolecules.

2 Methods

2.1 Learning $f_{structure}$ with a 3D-WGAN

As Step 1 of our procedure, we construct a generative model $G : \mathbb{R}^{100} \to \mathcal{S}_{pseudo}$ to learn the structural distribution of macromolecules $f_{structure}$. Our generative model combines the 3D-GAN [20] with the Improved Wasserstein GAN [8]. In our paper, we refer to our model as 3D Deep Convolutional Wasserstein GAN (3D-WGAN)[1]. The network architecture of our

[1]Background details about GAN, as well as specifics on the convolutional layer, batch normalization layer, activation layers and the minibatch discrimination layer can be found in Supplementary Section 1.4

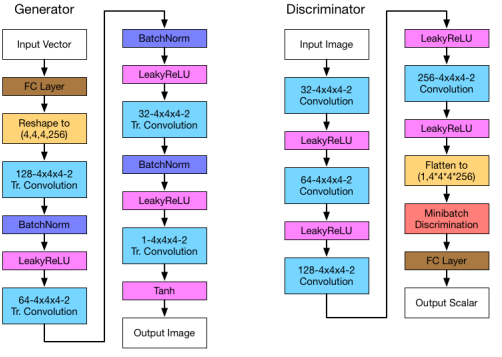

Figure 1: The network architecture for both the generator and discriminator of the 3D-WGAN. Each convolution layer is labeled in the format $N - K \times K \times K - S$, signifying N filters with kernel size K^3 and stride S.

3D-WGAN is presented in Figure 1. Inspired by [20], our network generator and the discriminator are implemented each with four convolution layers of stride 2 and kernel size 4^3, which we chose to be a factor of 64 to reduce the checkerboard artifact [16]. We found that using half as many filters as[20] was sufficient for the GAN to stabilize at producing good results. In the hidden layers, we used the LeakyReLU activation with $\alpha = 0.2$. Following [1, 8], we only used Batch Normalization [11] in the generator.

A common problem for GANs with especially limited data is mode collapse, which occurs when the generator only produces structures with very low diversity. We adopted the minibatch discrimination layer in the discriminator (see [18]) to reduce the collapse of the generator by penalizing low-entropy generators. This layer allows the discriminator to observe many samples at once, so that it can also take entropy of a batch of samples into account when deciding between real and pseudo images.

In our experiments, each macromolecular complex was represented as a 3D gray-scale image (a.k.a. density map) with 64^3 voxels and 0.6nm pixel size. We constructed our dataset with 15 experimental macromolecular complexes that are diverse in shape and size (see third row of Figure 3). To prevent overfitting to a specific orientation or structure, we performed data augmentation and rotated each structure 600 times for a total of 9000 training structures. As the 3D convolution operation is not rotation invariant, this data augmentation improves the training of the 3D-WGAN. We trained the 3D-WGAN using a batch size of 64 and the Adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.99$, and the learning rate as 0.0001, shown to be successful in previous works [17, 20]. Following [8], we trained the discriminator ten times as often as the generator, and we used a gradient penalty of 10.

2.2 Sampling pseudo-complexes far away from $C_{Interest}$

Under the null, the instance of $C_{Interest}$ should be not be covered by the distribution of pseudo-complexes. This procedure also reduces the chance of sampling pseudo-complexes so similar to $C_{Interest}$ that they could be viewed as copies of $C_{Interest}$ in the hypothesis test. With the following three-step procedure, we could sample pseudo-complexes from $f_{structure}$ to be "far away" from $C_{Interest}$ in the latent representation of $f_{structure}$:

1. Regressor (a.k.a. Inverse Generator): We trained a regressor for the inverse map of the generator, using a 3D extension of the AutoEncoder GAN model (see [14]). Instead of using cross-entropy loss, we used the sum squared error of the reconstructed images since our images were not normalized to $(0,1)$ by a sigmoid. The network architecture of the regressor is an exact mirror image of the 3D-WGAN generator.
2. Kernel Density Estimator (KDE): We trained a KDE to learn a distribution $\mathcal{E}$ of the latent representation of $C_{Interest}$, given by regressor's output for 300 random rotations of $C_{Interest}$. The KDE's bandwidth was determined using 3-fold cross validation.
3. Bayes Classifier: Let π ($\pi \ll 1$) be the prior probability of pseudo-complexes, which can be estimated from the data. Denote $\mathcal{N}$ as the standard 100-dimensional multivariate Gaussian. The decision boundary to distinguish the distribution of pseudo-complexes and $C_{Interest}$ is where the probability of a complex being a pseudo-complex is the same as the probability of a complex being the known complex $C_{Interest}$.

   According to Bayes rule, the rejection region can be written as:

$$\mathcal{R} = \{G(v) : v \in \mathbb{R}^{100} \text{ such that } \mathcal{N}(v) < \pi \cdot \mathcal{E}(v)\}$$

   . When we sample v from $\mathcal{N}$, we reject the members of $\mathcal{R}$.

2.3 Monte Carlo approach for evaluating the statistical credibility of template matching

After determining the complex of interest $C_{Interest}$ in Step 2, we perform a statistical assessment of template matching for Step 3 by calculating an empirical p-value using pseudo-macromolecular complexes. The true p-value p is the probability of obtaining results at least as extreme as the observed $c(P,T(C_{Interest}))$ given the null hypothesis H_0 is true. Since the distribution of C_0 under $f_{structure}$ is unknown, we use a Monte Carlo simulation to obtain an unbiased empirical p-value $\widehat{p}$ by ranking the observed test statistic $c(P,T(C_{Interest}))$ amongst the alignment scores of complexes sampled from the learned distribution of the 3D-WGAN. By the strong law of large numbers, our empirical p-value converges almost surely to the true p-value as the number of Monte Carlo samples $B \to \infty$.

$$p = \mathbb{E}_{H_0; C_0 \sim f_{structure}}[\mathbb{I}\{c(P,T(C_{Interest})) \le c(P,T(C_0))\} | C_0 \notin \mathcal{R}] \tag{3}$$

$$= \Pr(c(P,T(C_{Interest})) \le c(P,T(C_0)) | C_0 \notin \mathcal{R}) \tag{4}$$

$$\widehat{p} = B^{-1} \sum_{b=1}^{B} \mathbb{I}[c(P,T(C_{Interest})) \le c(P,T(C_0^{(b)}))] \xrightarrow[B\to\infty]{a.s.} p \tag{5}$$

where $\{C_0^{(b)}\}_{1 \le b \le B}$ are Monte Carlo samples from $f_{structure}$ and out of the rejection region $\mathcal{R}$, specified in Section 2.2.

The statistical credibility of the decision is measured by the p-value: the smaller the p-value, the more statistical credibility we have to reject the null and to support the alternative.

Therefore, when $c(P, T(C_{Interest}))$ is ranked in the highest 1% of the templates used for template matching (i.e. $\widehat{p} \leq 0.01$), we could conclude that P contains a macromolecule identical to $C_{Interest}$ with high confidence. We note here that in CECT template matching tasks, there is usually a large number of hypothesis tests for different subtomograms matching, which will consequently generate a large number of p-values. Multiple comparison is also suggested to be adjusted to control false discovery rate [3] to provide a stringent statistical criteria.

3 Results

3.1 Examples of pseudo-macromolecular complexes

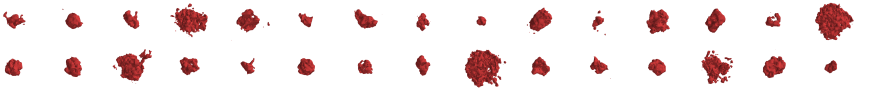

Figure 2: Random pseudo-macromolecular complexes generated with the 3D-WGAN.

Randomly selected pseudo-complexes from the 3D-WGAN are shown in Figure 2. Figure 3A shows the nearest neighbors of pseudo-complexes from $\mathcal{S}_{known}$. Most pseudo-complexes exhibit similar structure as their nearest neighbor in $\mathcal{S}_{known}$ and resemble the same macromolecular complex. Figure 3B shows the nearest neighbors of known complexes (labeled by their Protein Data Bank (PDB) ID[2]) from a set of 10,000 generated pseudo-complexes. We defined the metric between complexes as the L_2 norm on the fully-connected layer of the discriminator, which is a high-level feature representation of each complex [20]. Even as nearest neighbors of known complexes, the pseudo-complexes in Figure 3 are visually not identical to the known complexes. Non-rigid differences between the pseudo-complexes and the known complexes make these pseudo-complexes good candidates for hypothesis testing. This shows that our model can produce meaningful pseudo-complexes that have recognizable structural similarities to the training structures. A practical advantage to using the 3D-WGAN is that the generation of pseudo-complexes can be significantly sped up with a GPU. For example, our GTX 1080 Ti GPU could generate 10,000 structures in about 20s (0.002s/image).

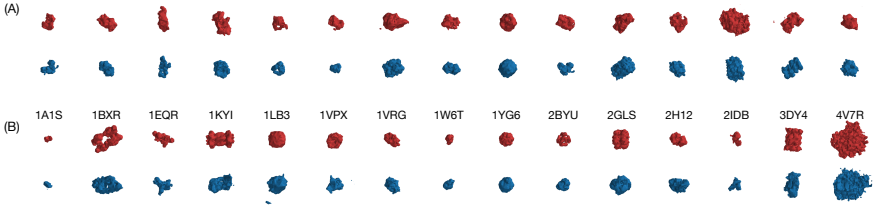

Figure 3: Let C_{ij} denote the structure in the i^{th} row and the j^{th} column. (A) Each (red) $C_{1j} \in \mathcal{S}_{pseudo}$, and (blue) $C_{2j} \in \mathcal{S}_{known}$ is the most similar structure to C_{1j} in $\mathcal{S}_{known}$. (B) Similarly, Each (red) $C_{3j} \in \mathcal{S}_{known}$ and is listed according to its PDB ID. Each $C_{4j} \in \mathcal{S}_{pseudo}$ is the most similar structure to C_{3j} in $\mathcal{S}_{pseudo}$.

[2] A table that associates PDB IDs to macromolecular complexes can be found in Supplementary Section 1.3

3.2 Statistical assessment of template matching on simulated subtomograms

Template matching is applied to real data to decide whether an unidentified subtomogram contains a macromolecule of identical structure to some known macromolecule. For every $KC \in \mathcal{S}_{known}$, we simulated[3] a subtomogram P_{KC} containing KC. The hypothesis test consists of 985 randomly generated pseudo-templates $T(C_0) \in \mathcal{S}_{pseudo}$ and 15 templates of known structure $T(KC) \in \mathcal{S}_{known}$. We performed template matching on P_{KC} following the three-step process described in Section 1. A simulated test was declared successful if both conditions are satisfied:

(Cond. 1) Highest correlation: KC is chosen to be $C_{Interest}$. In other words, KC satisfies $c(P_{KC}, T(KC)) \geq c(P_{KC}, T(C')), \forall C' \in \mathcal{S}_{known}$.

(Cond. 2) Low p-value: The hypothesis test has empirical p-value $\widehat{p} \leq 0.01$.

As we perform a test for each $KC \in \mathcal{S}_{known}$, we performed a total of 15 tests and achieved an average success rate of $12/15 = 80\%$, indicating high power of our hypothesis testing procedure. Two successful cases are shown in Figure 4. Others are shown in Supplementary Sections 1.5 (two unsuccessful cases) and 1.6 (rest of results).

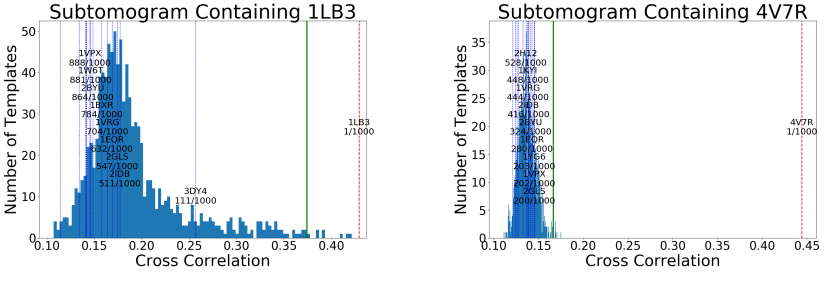

Figure 4: Results from successful hypothesis tests on simulated subtomograms containing 1LB3 and 4V7R. The blue histograms model the distribution of $c(P_{KC}, T(C_0)), C_0 \in \mathcal{S}_{pseudo}$. The vertical dashed lines denote $c(P_{KC}, T(C')), C' \in \mathcal{S}_{known}$, with the ten highest labeled with the PDB ID of C' and a rank out of the 1000 (i.e. the empirical p-value). $C_{Interest}$ is highlighted with the red color. The green line marks the p-value threshold of 0.01.

From the smooth distribution of histograms in Figure 4, we can deduce that our generative model samples diverse and realistic pseudo-complexes from $f_{structure}$, instead of reproducing the known complexes, which would result in the histograms clustering only around the vertical lines.

3.3 Statistical assessment of template matching on experimental subtomograms of ribosomes

With the same hypothesis testing procedure, we performed template matching on experimental subtomograms of Yeast 80S ribosome (PDB ID: 4V7R) from the EMPIAR-10045 database [12]. We used the 07 tomogram, which contains 376 subtomograms in total. The

[3] Details on simulation and template construction are described in Supplementary Section 1.2

experimental subtomograms were originally 200^3 voxels with 0.217nm pixel size. For our experiments, they were preprocessed with a 2.17nm Gaussian blur and resized to 64^3 voxels with 0.6nm pixel size, the same dimension and pixel size as the density maps of the training structures and the simulated subtomograms. As shown in Figure 5A, these subtomograms are visually much noisier than the simulated ones, which illustrates the fact that CECT template matching is extremely challenging. Without preprocessing, it is extremely difficult to even visually detect the macromolecular structure contained within the subtomograms.

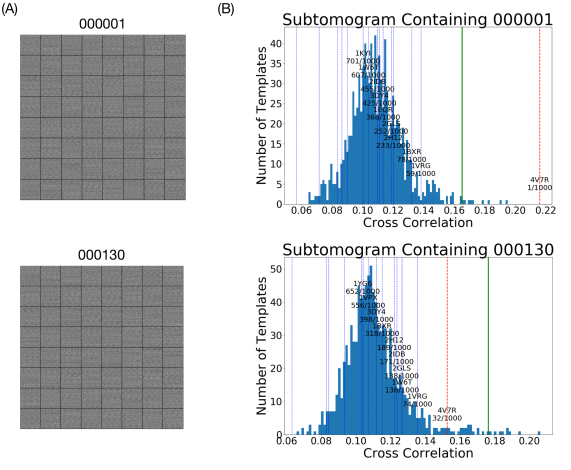

Figure 5: (A) Slices of raw subtomograms containing ribosome; (B) Results from hypothesis tests on experimental subtomograms of ribosome. 0000001 is successful and 000130 is unsuccessful.

With a p-value cutoff at 0.01, there were 264 (70.21%) successful tests, where Cond. 1 and Cond. 2 were satisfied for a known ribosome complex of interest. 45 subtomograms (11.97%) resulted in Cond. 1 failure. 67 subtomograms (17.82%) resulted in Cond. 2 failure. Figure 5B shows one successful and a case of Cond. 2 failure from our experiments.

We note that due to our rescaling procedure, some subtomograms may be missing part of their ribosome structure, which may cause template matching to fail. Therefore, we further selected a subset of 100 subtomograms with the highest ribosome alignment score, corresponding to high confidence in containing the whole ribosome structure. 92 tests were successful, 1 resulted in Cond. 1 failure, and 7 resulted in Cond. 2 failure.

3.4 Detecting false positives from template matching

When the subtomogram does not contain a macromolecule or when the macromolecule in the subtomogram does not match with any of the known templates, it is desirable for template matching to conclude that no template matches with the given subtomogram. This is not possible if the template with the highest alignment scores is always concluded to be a match. Even when thresholding, it is difficult to choose a single cutoff that works in all possible cases and requires hyperparameter tuning. Using 20 experimental subtomograms that do not contain any macromolecules, we performed template matching along with our hypothesis testing procedure and we were able to prevent 8 (40%) of the false positives that would have occurred if we simply chose the highest alignment score as a match. Our statistical testing

method makes more reliable claims on the results of template matching and can significantly reduce the number of false positives.

3.5 Learning the shape manifold of macromolecules

Similar to previous works [17, 20] with GANs, we found that our model was able to capture the shape manifold of macromolecular complexes. We show this by interpolating between the latent vectors of S and D, resulting in a smooth transition from a proteasome (PDB ID: 3DY4) to a ribosome (PDB ID: 4V7R) as shown in Figure 6. Our starting point S was the latent vector of the nearest pseudo-complex for the proteasome, and our ending point D was the latent vector of the nearest pseudo-complex for the ribosome. We generated the pseudo-complex at the i^{th} step with the input vector $S + (D - S) \cdot i$. This smooth "deformation" from one structure to another illustrates the shape manifold of macromolecules and results in an effect similar to deformable image registration. Since deformable image registration can be quite computationally expensive (i.e. Large Deformation Diffeomorphic Metric Mapping [15]), with a well-trained model, our 3D-WGAN model could potentially be used as a computationally efficient heuristic for deformable registration of 3D shapes.

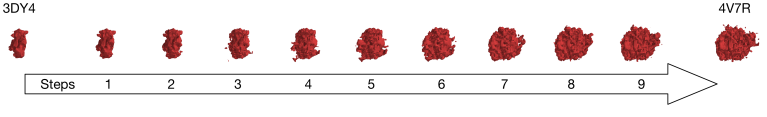

Figure 6: Intermediate shapes generated by interpolating between the latent representation of proteasome (PDB ID: 3DY4) and ribosome (PDB ID: 4V7R). This shows that our generative model can learn the shape manifold of macromolecular structures.

4 Conclusion

Without hypothesis testing, existing template matching approaches are not rigorous and are not statistically credible enough. Physical limitations to CECT remain a major difficulty that can bias the cross-correlation score and may even cause template matching to fail. To reliably conclude the identity of a subtomogram and to reduce false positive rates, we propose an image-derived approach for performing hypothesis testing for template matching by constructing a generative model for macromolecular complexes. We used the 3D-WGAN since it could efficiently produce a diverse population of novel pseudo-macromolecular complexes that are not simply rigid rotations of the original training structures. By sampling from the learned distribution of macromolecules $f_{structure}$, we could successfully conclude, with both simulated and experimental subtomograms, with high statistical credibility that the given subtomogram contains a macromolecule identical to the complex of interest. In addition, the 3D-WGAN generative model has potentially other applications in CECT. We have shown that the 3D-WGAN can learn the shape manifold of macromolecules [20]. By interpolating between latent representations, a 3D-WGAN can be used to visualize smooth transitions between structures. This ability for smooth deformations can be potentially extended to a heuristic for deformable image registration.

In addition, our approach is not limited to the 3D-WGAN and works with any reasonable generative model. A future work would include experimenting other generative approaches

(e.g. VAE, shape-space modeling [15]) with the hypothesis testing procedure. The current GAN approach may also be improved with larger training sets or more complex architecture (e.g. [13]). Finally, our procedure of using generative modeling, particularly the GAN, for hypothesis testing can be generalized to be applicable in many other template matching tasks with 2D or 3D images. Some prominent examples include object detection and fingerprint scans, where statistical tests are very important to ensure the quality of the result and to avoid false positives [19].

5 Acknowledgements

We thank Dr. Robert Murphy and Dr. Christopher Langmead for suggestions. This work was supported in part by U.S. National Institutes of Health (NIH) grant P41 GM103712. MX acknowledge support from Samuel and Emma Winters Foundation. EX acknowledges supports of NIH Grant R01GM114311, NIH Grant P30DA035778, and Pennsylvania Department of Health CURE Grant BD4BH4100070287.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[2] Shoh Asano, Benjamin D Engel, and Wolfgang Baumeister. In situ cryo-electron tomography: a post-reductionist approach to structural biology. *Journal of molecular biology*, 428(2):332–343, 2016.

[3] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[4] Jochen Böhm, Achilleas S Frangakis, Reiner Hegerl, Stephan Nickell, Dieter Typke, and Wolfgang Baumeister. Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences*, 97(26):14245–14250, 2000.

[5] F. Förster, S. Pruggnaller, A. Seybert, and A.S. Frangakis. Classification of cryo-electron sub-tomograms using constrained correlation. *Journal of structural biology*, 161(3):276–286, 2008. ISSN 1047-8477.

[6] Zachary Frazier, Min Xu, and Frank Alber. Tomominer and tomominercloud: A software platform for large-scale subtomogram structural analysis. *Structure*, 25(6):951–961, 2017.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017.

[9] Bong-Gyoon Han, Ming Dong, Haichuan Liu, Lauren Camp, Jil Geller, Mary Singer, Terry C Hazen, Megan Choi, H Ewa Witkowska, David A Ball, et al. Survey of large protein complexes in d. vulgaris reveals great structural diversity. *Proceedings of the National Academy of Sciences*, 106(39):16580–16585, 2009.

[10] Adery CA Hope. A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 582–598, 1968.

[11] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[12] Andrii Iudin, Paul K Korir, José Salavert-Torres, Gerard J Kleywegt, and Ardan Patwardhan. Empiar: a public archive for raw electron microscopy image data. *Nature methods*, 13(5):387–388, 2016.

[13] Tero Karras, Tima Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. 2017.

[14] Junyu Luo, Yong Xu, Chenwei Tang, and Jiancheng Lv. Learning inverse mapping by autoencoder based generative adversarial nets. In Derong Liu, Shengli Xie, Yuanqing Li, Dongbin Zhao, and El-Sayed M. El-Alfy, editors, *Neural Information Processing*, pages 207–216, Cham, 2017. Springer International Publishing. ISBN 978-3-319-70096-0.

[15] Robert F Murphy. Building cell models and simulations from microscope images. *Methods*, 96:33–39, 2016.

[16] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016. doi: 10.23915/distill.00003. URL http://distill.pub/2016/deconv-checkerboard.

[17] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[18] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.

[19] Paridhi Swaroop and Neelam Sharma. An overview of various template matching methodologies in image processing. *International Journal of Computer Applications*, 153(10), 2016.

[20] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems*, pages 82–90, 2016.

[21] M. Xu, M. Beck, and F. Alber. High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching. *Journal of Structural Biology*, 178(2):152–164, 2012.

[22] Min Xu, Xiaoqi Chai, Hariank Muthakana, Xiaodan Liang, Ge Yang, Tzviya Zeev-Ben-Mordehai, and Eric Xing. Deep learning based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *ISMB/ECCB 2017, Bioinformatics (in press); arXiv preprint arXiv:1701.08404*, 2017. doi: 10.1093/bioinformatics/btx230.