Eurographics Workshop on 3D Object Retrieval (2021)
S. Biasotti, R. M. Dyke, Y. Lai, P. L. Rosin, and R C. Veltkamp (Editors)

Short Paper

SHREC 2021: Surface-based protein domains retrieval

F. Langenfeld[1], T. Aderinwale[2], C. Christoffer[2], W.–H. Shin[3], G. Terashi[4], X. Wang[2], D. Kihara[2,4], H. Benhabiles[5], K. Hammoudi[6,7], A. Cabani[8], F. Windal[5], M. Melkemi[6,7], E. Otu[9], R. Zwiggelaar[9], D. Hunter[9], Y. Liu[10], L. Sirugue[1], H.–N. H. Nguyen[11,12], T.–D. H. Nguyen[11,12], V.–T. Nguyen–Truong[11,12], D. Le[11,12], H.–D. Nguyen[11,12], M.–T. Tran[11,12,13], M. Montès[1]

[1]GBCM, EA 7528, Conservatoire National des Arts-et-Métiers, HESAM Université, Paris, France
[2]Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA
[3]Department of Chemical Science Education, Sunchon National University, Suncheon 57922, Republic of Korea
[4]Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA
[5]Univ. Lille, CNRS, Centrale Lille, Univ. Polytechnique Hauts-de-France, Junia, UMR 8520 - IEMN, F-59000 Lille, France
[6]Université de Haute-Alsace, Department of Computer Science, IRIMAS, F-68100 Mulhouse, France
[7]Université de Strasbourg, France
[8]Normandie University, UNIROUEN, ESIGELEC, IRSEEM, 76000 Rouen, France
[9]Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3FL, UK
[10]Department of Computer Science, Edge Hill University, Ormskirk, L39 4QP, UK
[11]University of Science, VNU-HCM, Vietnam
[12]Vietnam National University, Ho Chi Minh City, Vietnam
[13]John von Neumann Institute, VNU-HCM, Vietnam

Abstract

Proteins are essential to nearly all cellular mechanism, and often interact through their surface with other cell molecules, such as proteins and ligands. The evolution generates plenty of different proteins, with unique abilities, but also proteins with related functions hence surface, which is therefore of primary importance for their activity. In the present work, we assess the ability of five methods to retrieve similar protein surfaces, using either their shape only (3D meshes), or their shape and the electrostatic potential at their surface, an important surface property. Five different groups participated in this challenge using the shape only, and one group extended its pre-existing algorithm to handle the electrostatic potential. The results reveal both the ability of the methods to detect related proteins and their difficulties to distinguish between topologically related proteins.

CCS Concepts

• ***Applied computing*** → ***Computational biology;*** • ***General and reference*** → ***Evaluation;***

1. Introduction

Proteins are linear assemblies of amino-acids that fold in specific, energy-driven 3D structures [Kar11] that are linked to their biological activity. Identifying similarities within protein structures is therefore of tremendous importance in various fields, from biochemistry to drug design. Numerous methods, based on the analysis of the 3D point clouds defined by the 3D coordinates of the protein atoms [Zha05], rely mostly on the conserved core structure of proteins and may be inefficient to detect proteins sharing surface similarity. The protein surface is a higher-level description of the protein structure that abstracts the underlying protein sequence, structure and fold into a continuous shape with geometric and chemical features that fingerprint its interactions with the other molecules of its environment [SPNW04]. Only a limited number of methods have been proposed so far [SLL*08b, GSM*19, ZSSZ20]. The aim of this challenge is to assess the performance of five currently available methods to retrieve similar proteins using the 3D meshes describing their surfaces (shape-only challenge), or using both the surface and the electrostatic potential at the surface (shape+electrostatics challenge).

2. The Dataset

2.1. Constitution of the SHREC'21 dataset

The dataset relies on the Pfam 33.1 database [MCW*20], which classifies protein sequences into domains and families. Protein domains of structures from the Protein Data Bank (PDB, [BHN03]) can therefore be attributed to a Pfam domain. To build up the challenge dataset, we selected 10 Pfam domains with a large range of fold similarity (from 0.19 to 0.84, Fig 1): SH2, SH3, SH3_2, PDZ, PDZ_6, m50, bromo–domain, DNA-binding domain of STAT protein, PHD-finger and C2H2 Zinc-finger. TM-scores below 0.17 correspond to unrelated proteins, while those above 0.5 usually indicate two structures having the same fold [ZS04]; compared to previous SHREC challenges, this dataset therefore contains a greater

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

DOI: 10.2312/3dor.20211308

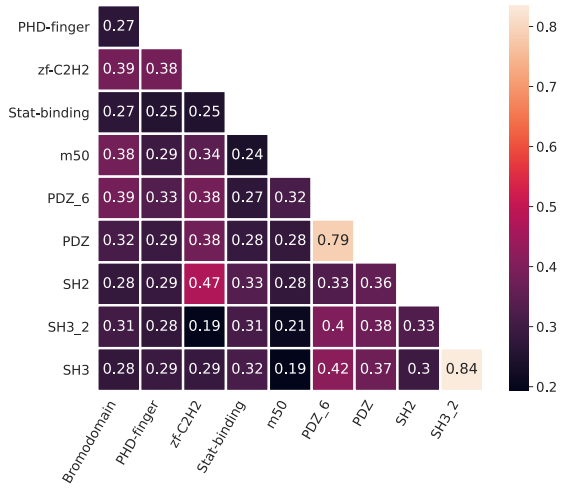

Figure 1: *Structural similarity between the protein structure queries. The TM-score (in the (0, 1] range) measures the topological similarity between two protein structures: the higher the TM-score, the more similar the two structures.*

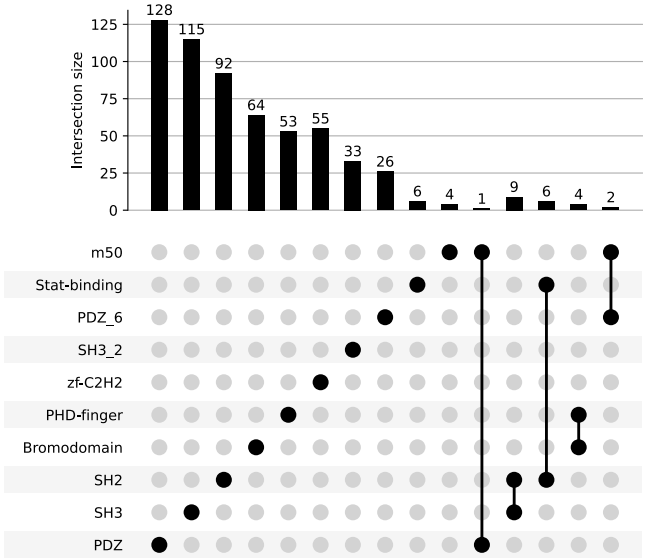

Figure 2: *Upset plot of the dataset structure. The dataset is composed of 554 individual shapes, of which 22 bears two of the domains of the dataset.*

amount of structures with intermediate structural similarity (*i.e.* with TM-scores from 0.17 to 0.5).

For each selected domain, all corresponding structures from the PDB were listed, and the best resolution structures were retrieved to serve as a query for the challenge. For some of these domains, only a limited number of PDB structure is available. The remaining structures were filtered according to their Uniprot [Con20] identifier, duplicates were discarded and only the best resolution structures for each Uniprot entry were kept. To be noted, each structure may contain more than one domain (see Figure 2 for the final structure of the dataset).

The solvent-excluded surface of all protonated structures were computed using the default parameters of EDTSurf [XZ09], discarding inner cavities. We computed the electrostatics using APBS suite [BSJ*01], and used the *multivalue* software to compute the electrostatic potential at the mesh vertices locations. The two datasets proposed (shape-only and shape+electrostatic) includes 554 molecular surfaces which were eventually made available to the challenge participants, along with the 2 sets of 10 queries. Figure 1 represents the TM-scores matrix for all queries of the dataset.

2.2. Challenge proposed to the participants

The participants were asked, given each of the query surfaces, to retrieve the molecular surfaces of proteins that encompass the same domain as the query. Each query-to-dataset-surface distance was expected to be expressed as a dissimilarity score. Each participant was allowed to submit one dissimilarity matrix for each dataset: one matrix for the shape-only dataset, and one matrix for the shape+elecrostatic dataset.

3. Participants and methods

Among the seven groups that initially registered to this challenge, only 5 were able to produce the results in time and returned a shape-only dissimilarity matrix. Most of the participants develop methods dedicated to the analysis of 3D surfaces and are not accustomed to the use of additional data, only one method (3DZD) has been adapted in time to handle the shape+electrostatics.

3.1. Network trained with encoded 3DZD (3DZD) by T. Aderinwale, C. Christoffer, W.–H. Shin, G. Terashi, X. Wang & D. Kihara

This group submitted two (shape-only and shape+electrostatic) dissimilarity matrices of the target proteins to the 10 query proteins provided by the organizers. These methods are based on the 3D Zernike Descriptor (3DZD). 3DZD is the rotation-invariant shape descriptor derived from the coefficients of 3D Zernike-Canterakis polynomials [Can99].

Similar to SHREC'20 [LPL*20], this group trained two types of neural network to output a score that measures the (dis)similarity between a pair of protein shapes. Briefly, the first framework (the Extractor model) was structured into multiple layers: an encoder layer with 3 hidden units of size 250, 200 and 150, a feature comparator layer which computes the Euclidean distance, cosine distance, element-wise absolute difference and product, and a fully connected layer with 2 hidden units of size 100 and 50. There were multiple hidden units in each layer, and this group used the ReLu activation function in all except the output of the fully connected layer where the sigmoid activation function was used to output the probability that the two proteins belong to the same protein- or species-level in the SCOPe dataset classification [FBC13].

The network was trained on the latest SCOPe dataset of 259,385

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

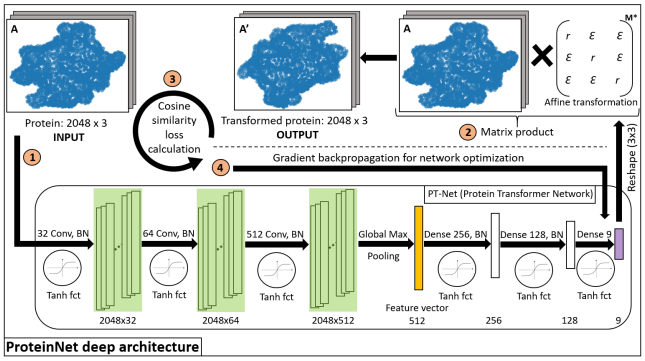

Figure 3: *ProteinNet deep architecture for protein point cloud transformation into canonical representation.*

protein structures. 2,500 protein structures were set aside for network validation. Proteins in Class I (Artifacts) were removed. Each of the two network frameworks were trained with two datasets. The first dataset was 3DZDs of surface shape of proteins and the second one was feature vectors that concatenate 3DZD of shape and 3DZD of the electrostatic properties.

This group examined the performance of the networks on the validation dataset to determine which models to use. For the shape-only dataset, the 3DZD group submitted predictions generated by the Extractor model. For the shape+electrostatic dataset, this group submitted the average predictions between the Extractor model and the End-to-end model.

For each protein in the provided dataset, the 3DZD group performed a pre-processing step as follows: (1) The PLY mesh data file was converted to a volumetric skin representation (Situs file) where points within 1.7 grid intervals were assigned with values that were interpolated from the mesh [SLL*08b]. For the electrostatic features, the interpolated values were the potentials at the mesh vertices. For the shape feature, a constant of 1 was assigned to grids that overlap with the surface. (2) The resulting Situs file was fed into the EM-Surfer pipeline [ERXH*15] to compute 3DZD.

3.2. ProteinNet: Deep learning based protein characterization from 3D point clouds (ProteinNet) by H. Benhabiles, K. Hammoudi, A. Cabani, F. Windal & M. Melkemi

This group proposes a deep learning approach to calculate a protein descriptor from its 3D point cloud. To this end, the ProteinNet group developed a variant of PointNet [QSMG16] which is a point cloud deep architecture dedicated for 3D classification and segmentation. This group adapted this architecture in order to learn an affine transformation matrix that allows to align the coordinates of the input 3D protein point cloud into a canonical representation. The new representation maintains interesting properties demonstrated in [QSMG16], including invariance to rigid geometric transformations as well as point order permutations. The ProteinNet deep architecture is illustrated in Figure 3. More specifically, the architecture is based on a PT-Net module (Protein Transformer Network) which is inspired from the T-Net (Transformer Network) module of the original PointNet architecture. The PT-Net module is trained to predict an affine transformation matrix M that is constrained to be close to an orthogonal matrix, namely $|(M.M^t) - I| = 0$ (step 1 in Figure 3). The matrix M is used to transform the input protein into its canonical representation (step 2 in Figure 3). A cosine similarity loss between the original protein and the transformed one is then calculated (step 3 in Figure 3) in order to back-propagate the error over the network (step 4 in Figure 3) and optimize the matrix M.

PT-Net module The module is composed of a sequence of 3 convolution blocks (32, 64 and 512 layers) followed by a global max pooling layer and 3 successive dense layers (256, 128 and 9). As shown in Figure 3, each convolution block as well as the dense layers (except the last one) undergo a batch normalization and a tangent hyperbolic activation function. The last dense layer of 9 units is reshaped to output the (3×3) M matrix.

Data preparation and architecture training All the proteins of the dataset of the challenge have been sampled to 2,048 points using a Poisson disk sampling technique [Yuk15] and then normalized into a zero-center unit sphere based on their respective minimum bounding spheres [BHW*19]. The architecture has then been trained using a batch size of 16 on 80% of the dataset over 150 epochs and validated on the remaining 20% of the data. The training data were augmented on-the-fly (during the training process) by adding some geometric noise.

Protein feature extractor The trained ProteinNet model has then been exploited to calculate a protein feature descriptor, for each input protein, by extracting its intermediate Global Max Pooling hidden layer. This descriptor corresponds to a 1-dimension vector of 512 values.

Dissimilarity matrix computation The dissimilarity matrix between the ten protein shape queries and the set of 554 protein shapes has been calculated using Euclidean distance between their respective 512 feature vectors.

3.3. Agglomeration of local Augmented Point-pair Feature Descriptors with Fisher Kernel and Gaussian Mixture Model (APPFD-FK-GMM) by E. Otu, R. Zwiggelaar, D. Hunter & Y. Liu

This section presents a novel APPFD-FK-GMM 3D shape retrieval method based on Fisher Kernel (FK) and Gaussian Mixture Model (GMM) agglomeration of the Augmented Point-pair Feature Descriptor (APPFD) [OZHL19]: a 3D key point shape descriptor that robustly captures the physical geometric characteristics of 3D surface regions. Previous APPFD binning technique involves bucketting each of the 6-dimensional features of the APPFD into a multi-dimensional histogram with at least 7 bins in each feature-dimension, resulting to approximately $7^6 = 117649$-dimensional final feature-vector (APPFD), which is very high-dimensional final descriptor. In this work, we contribute a simpler approach, where each of the 6-dimensional feature is binned into a 1-dimensional histogram with 35 bins for each feature-dimension to produce a 210-dimensional local descriptor (APPFD) for every key point or

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

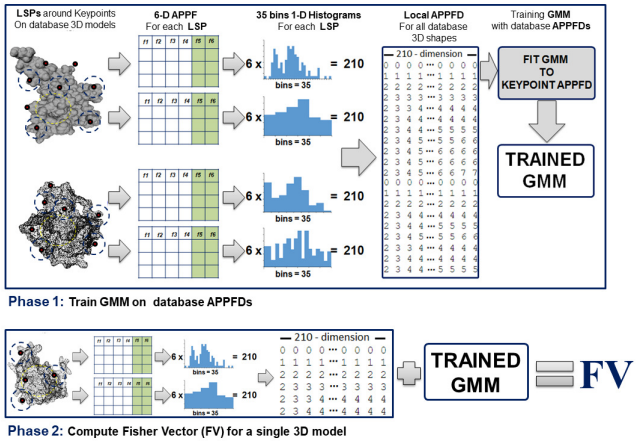

Figure 4: *APPFD-FK-GMM processing pipeline.*

local surface patch (LSP). Finally, the locally computed APPFDs are agglomerated into a compact code called the Fisher Vector (FV) with 4210 dimension, which is L_2 and power-normalized, and represents a single protein model, using the FK and GMM.

This work contributes a simple, efficient, robust, and compact representation, describing the geometry of 3D protein surfaces, with a knowledge-based (i.e. non-learning) approach. While a single protein surface in this challenge contains an average of 120,000 vertices and 200,000 triangular faces, our implementation address this very high data-structure by reducing 3D protein surface representation to as low as 3,500 points sample.

The APPFD-FK-GMM Method This method involves two main stages: (i) Computing local APPFDs for selected key points, and (ii) Key points APPFDs aggregation with FV and GMM described below. Figure 4 shows the processing pipeline of the APPFD-FK-GMM algorithm, and the reader is refered to [TBG*20], for further details regarding this method.

Experimental Settings and Running Time Matching two APPFD-FK-GMM descriptors representing two different protein surfaces is done with the L_2 norm. We submit a $[10 \times 554]$ dissimilarity matrix D, where the entry $D = [i, j]$ corresponds to the L_2 distance from i^{th} *FV* in the *query* set to the j^{th} *FV* in the *collection* set.

3.4. Projected Wave Kernel Signature Maps (PWKSM) by L. Sirugue & M. Montès

This method is based on the 2D projection of the surface and the Wave Kernel Signature (WKS) descriptor. Wave Kernel Signature [ASC11] is an isometric invariant descriptor that has been extensively improved and used in the field of computer vision [RRBW*14,BMM*15,LW15,ZLL*18]. This group have combined WKS with a 2D projection on a unit sphere [AHTK99]. Lowering one dimension of the space allows us to have a fast and dense comparison of the surface while having a smaller storage size for files.

Descriptor calculation In a first step, the WKS descriptor is computed on the surface of the 3D object for each point of the mesh. The surface is flattened on the unit sphere using a conformal transformation [AHTK99]. Then, the 2D spherical coordinates of the unit sphere are converted into 2D cartesian coordinates on the plane [CLM17]. A maps of size $(\theta_{max} - \theta_{min})/\delta, (\phi_{max} - \phi_{min})/\delta$ is created. θ_{max} and θ_{min} are the maximum and minimum values of θ on the unit sphere and same with ϕ, each representing an angle coordinate. δ is a coefficient to adapt the resolution. This type of projection is similar to topographic maps, that is why this group called this descriptor Projected Wave Kernel Signature Maps (PWKSM). An interpolation in the space of discrete integers is done to densify the maps. To reduce impact of deformation at the poles when converting to 2D cartesian coordinates, the PWKSM group computed 7 differents maps with different pole orientations.

Comparison A dense comparison is made using GPGPU sum reduction technique. Each point of a PWKSM is compared to all points of another PWKSM. The Earth Mover's distance L is used to compare the WKS descriptor of each point. Then, the smallest distance between a point of a first PWKSM T and all points of a second PWKSM V is selected. The sum of all the smallest distances for each point of the first PWKSM are summed to create the score S_T. The same is done for computing S_V.

$$S_T(T,V) = \sum_{k_T=1}^{N_T} \min_{k_V} L(T(k_T), V(k_V)) \tag{1}$$

The final score is the average of S_T and S_V defined as follows :

$$S = \frac{S_T + S_V}{2} \tag{2}$$

3.5. Graph-based learning methods for Surface-based protein domains retrieval (DGCNN) by H.–N. H. Nguyen, T.–D. H. Nguyen, V.–T. Nguyen-Truong, D. Le, H.–D. Nguyen & M.–T. Tran

In this deep learning method, this group exploits the availability of protein class labels from [RFB*21] to optimize the representation of protein surfaces without any additional properties. Particularly, this group designed a message-passing graph convolutional neural network (MPGCNN) with the Edge Convolution (EdgeConv) paradigm [WSL*19] for the protein classification objective. Then, the latent representation of protein surfaces from this neural network is used for the retrieval task in this challenge.

Data preprocessing For the meshes in each 3D model of a protein surface, this group first sampled 512 points on the surfaces of the meshes based on the area of the meshes. Then, to re-assign the topological structures for sampled points, this group connected each nodes with their k-Nearest Neighbors based on their original coordinates ($k = 16$).

Edge Convolution In this geometry-only setting, the initial node features was the coordinates of sampled points. Each protein sur-

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

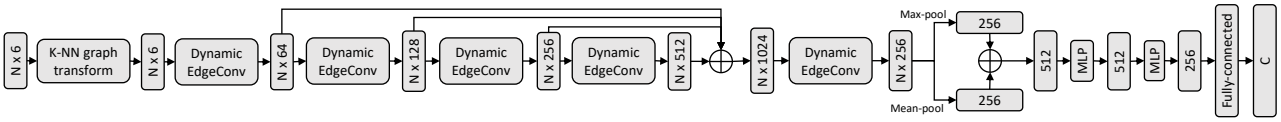

Figure 5: *Dynamic Edge Convolutional Neural Network*

face was represented by a k-Nearest Neighbors graph generated in the preprocessing step with 512 vertices (nodes).

The module that performed the graph message-passing function is the EdgeConv layer [WSL*19]. In the EdgeConv layer, the information of a vertex i after layer l is calculated as follows:

$$x_i^{l+1} = max_{j \in N}\, h(x_i^l, x_j^l) \tag{3}$$

where N is the neighboring vertices of vertex i with

$$h(x_i^l, x_j^l) = ReLU(MLP(x_i^l \oplus x_j^l)) \tag{4}$$

where ReLU is Rectified Linear Unit (in this implementation, the DGCNN group used LeakyReLU - a variant of ReLU), MLP is a standard multilayer perceptron (MLP), $\oplus$ is the concatenation operator.

In this implementation, the DGCNN group used a dynamic variant of EdgeConv instead of the standard EdgeConv described above. At each Dynamic EdgeConv layer, each vertex's k-Nearest Neighbors was re-calculated in the feature space produced by the previous layer, before applying the standard EdgeConv operation. After the graph was recomputed, standard EdgeConv operation was performed.

After the preprocessing phase, the vertex features first went through 4 layers of Dynamic EdgeConv. The dimensions of output features for each vertex after these first-4 layers were $64, 64, 128, 256$, respectively. Then, the outputs of these 4 layers were concatenated to become a 512-dimensional vector for each vertex. This 512-dimension vector was then fed through another Dynamic EdgeConv layer, creating the output vector with 512 dimensions v. The feature vector v was pooled using the concatenation of the outputs of a *max-pooling* and a *mean pooling* layer to generate the first graph-level feature vector. This vector was passed through two MLP blocks with BatchNorm, Leaky-ReLU, and Dropout layers. Finally, the vector was passed through a Fully-Connected layer for classification.

The latent representation of the graph was extracted as vectors by removing the last Fully-Connected layer from the network. The retrieval task was then performed by exploiting the $L2$-distances between these vectors.

4. Results

All teams returned a dissimilarity matrix for the shape-only dataset, and only one method (3DZD) was adapted to handle the shape+electrostatics dataset. We briefly present the corresponding results in this section.

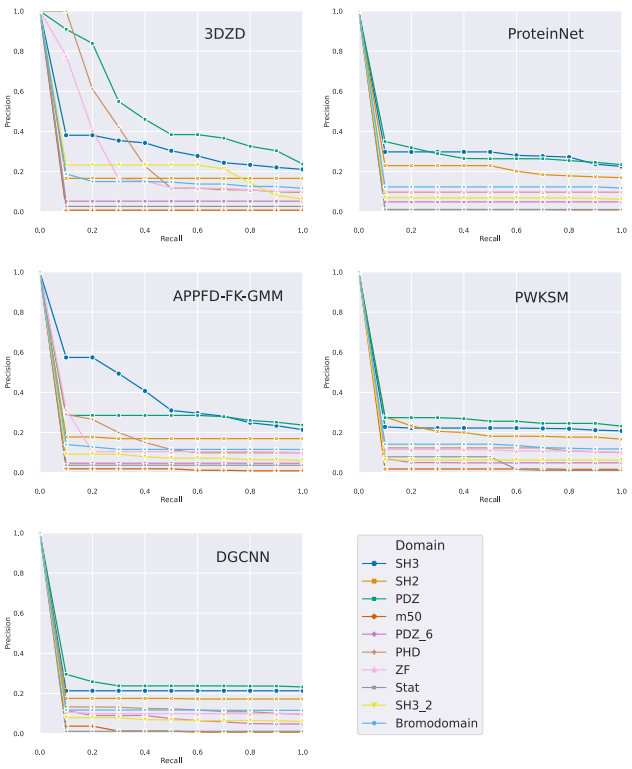

Figure 6: *Per-query precision-recall curves for the shape-only dataset, for each method. All plots are colored according to the legend on the bottom right of the figure.*

Method	Nearest Neighbor	First Tier	Second Tier	Mean Reciprocal Rank
3DZD	**0.5**	**0.160**	**0.292**	**0.523**
ProteinNet	0	0.088	0.195	0.126
APPFD	0.3	0.136	0.237	0.410
PWKSM	0.1	0.105	0.201	0.236
DGCNN	0.1	0.098	0.189	0.193

Table 1: *Summary of the evaluation metrics for the shape-only dataset.*

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

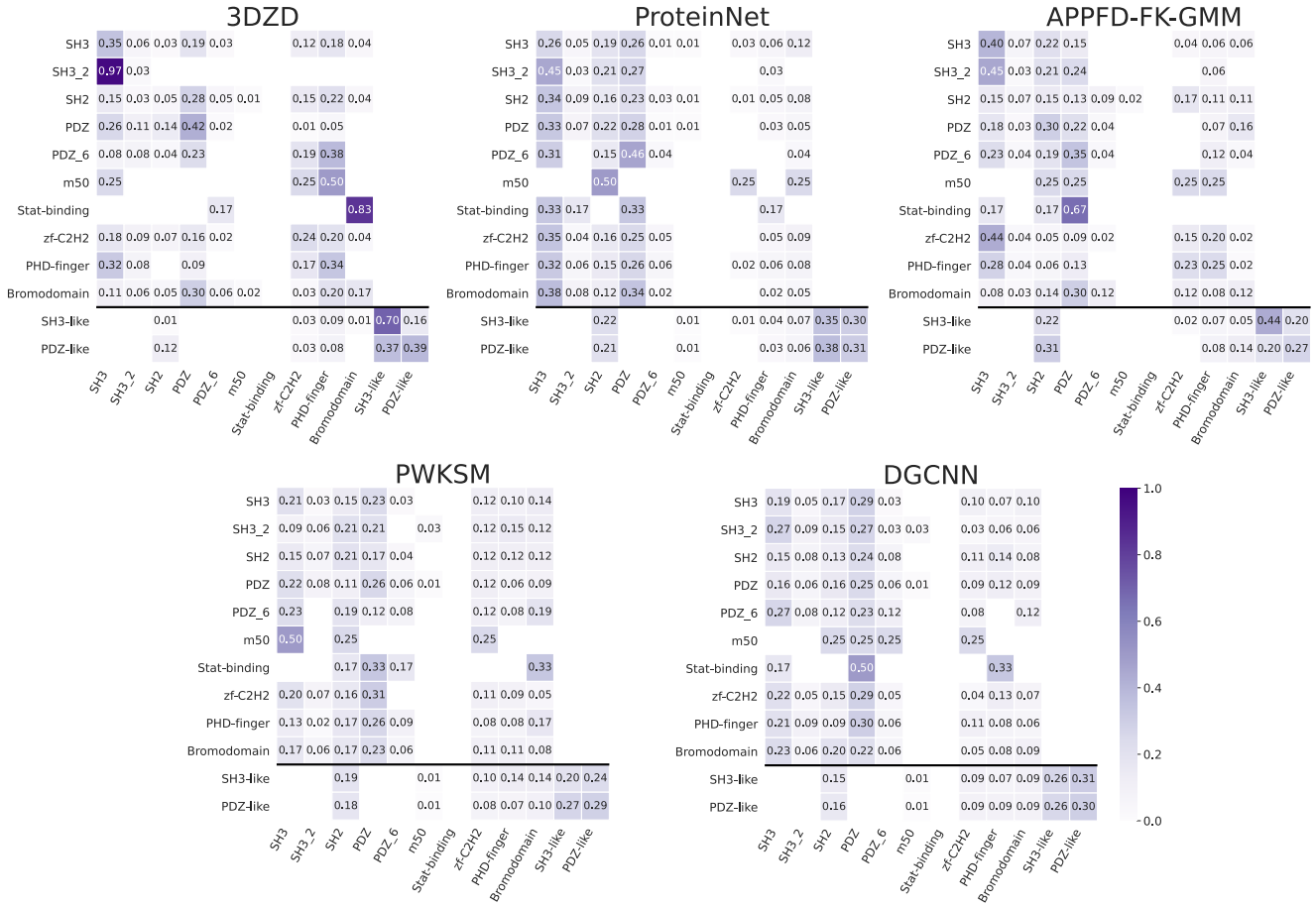

Figure 7: *Confusion matrices of all methods for the shape-only dataset. The color range is the same for all matrices. Confusion ranges from 0 (white background) to 1 (deep purple background).*

The best method (3DZD) achieved an overall level of 0.5 for the nearest neighbor metric, 0.160 for the first tier, 0.292 for the second tier and 0.523 for the mean reciprocal rank (Table 1). These results must be balanced by the fact that a few classes have only a small number of models (Figure 2). The precision-recall curves for each individual classes (Figure 6) show a quick drop of the precision at low recall values, except for a few exceptions (green curve, top left plot of Figure6, corresponding to the PDZ class for the 3DZD method, for instance) that display medium precision values at medium recall. The confusion matrices shown in Figure 7 combined with Figure 1 allow us to put the performance into perspective. For instance, PDZ and PDZ_6 domains are topologically very similar (TM-score: 0.79, Figure 1). When using the PDZ_6 query, ProteinNet retrieved only 1 (4%) of the 26 PDZ_6 shapes within the first 26 retrieved results, but also 12 (46%) shapes from the PDZ class (Figure 7, top right confusion matrix).

The evaluation metrics for the shape+electrostatics dataset are listed in Table 2, and show similar trends compared to the shape-only dataset. The precision-recall curves (Figure 8) show a similar overall behavior for the 3DZD method. T The confusion matrices (Figure 9) are in line with the previous results, indicating that 3DZD perform similarly in terms of overall performance but with a few differences at the per-class results.

Method	Nearest Neighbor	First Tier	Second Tier	Mean Reciprocal Rank
3DZD	**0.5**	**0.160**	**0.321**	**0.454**

Table 2: *Summary of the evaluation metrics for the shape+electrostatics dataset.*

5. Discussion and concluding remarks

The 3DZD method combines the use of 3D Zernike polynomials and a neural network trained on the SCOPe [FBC13] database, whose classification overlaps with the Pfam database [MCW*20] classification. The DGCNN used the data from another SHREC'21

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

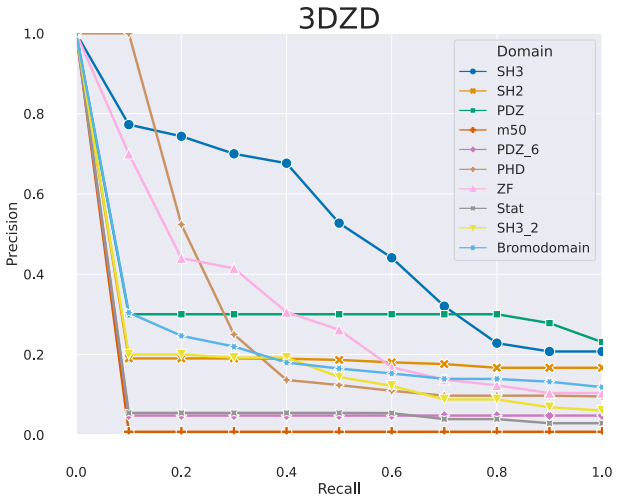

Figure 8: *Per-query precision-recall curves for the shape+electrostatics dataset, for each method.*

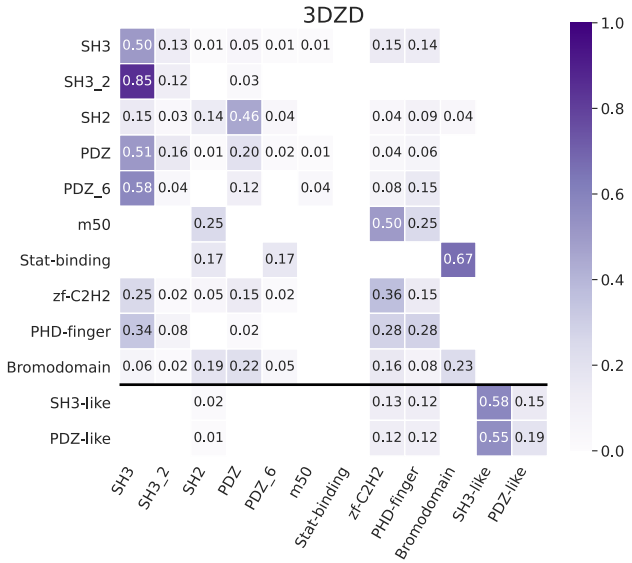

Figure 9: *Confusion matrices of all methods for the shape+electrostatics dataset. Confusion ranges from 0 (white background) to 1 (deep purple background).*

challenge [RFB*21], whose classification is also derived from the SCOPe database. The DGCNN and 3DZD methods were therefore trained on similar data, but results in different performance. DGCNN, ProteinNet and APPFD-FK-GMM methods down-sample the initial point clouds, among which the ProteinNet and APPFD-FK-GMM methods apply the more severe down-sampling. The APPFD-FK-GMM group, however, was able to better retrieve relevant results within the first hits (Table 1), indicating that down-sampling is not prohibited. As shown by the confusion matrices (Figure 7), a few queries were difficult to handle for all methods. Queries that are closely related (PDZ_6 / PDZ and SH3_2 / SH3 classes) result in results that highlight the difficulties of all methods to distinguish between closely related proteins using their shape only. Also, the DNA-binding domain of STAT proteins was seperated from the rest of the protein to serve as a query, while only a set of whole STAT proteins is present in the dataset. None method is able to retrieve a STAT protein within the first results using such a query. Methods dedicated to local similarity search may overcome this issue.

The results showed that the electrostatics only marginally improved the results with the 3DZD. These results are in line with [SLL*08a] which showed that electrostatics is best used to discriminate between very similar proteins. No general rule can be extracted, as only one group returned a shape+electrostatics matrix. Most groups could not extend their methods to handle electrostatics due to the time constraints.

Overall, this challenge revealed that satisfactory solutions exist to distinguish between loosely related proteins but also revealed some limits of these methods. Closely related proteins, *i.e.* proteins with a high topological similarity and limited changes of their amino-acid sequences, are still difficult to discriminate. The method using 3D Zernnike descriptor obtained the best overall results. Besides, this method is extended to handle additional data such as the electrostatics potential at the surface of the protein. This method, as well as all the other methods, are shared (see Section 3). In the future, the discrimination of closely related proteins based on their surfaces and / or their surficial properties could be a topic of a dedicated SHREC challenge, and a good indicator of the progress performed in this field.

Acknowledgments

The authors thank the 3DOR 2021 Workshop organizing committee for maintaining this workshop despite the current COVID-19 pandemic. F. Langenfeld, L. Sirugue and M. Montès are supported by the ERC Executive Agency (research grant number 640283). D. Kihara acknowledges supports from the National Institutes of Health (R01GM133840, R01GM123055) and the National Science Foundation (DBI2003635, CMMI1825941, and MCB1925643). C. Christoffer is supported by NIGMS-funded predoctoral fellowship (T32 GM132024).

References

[AHTK99] ANGENENT S., HAKER S., TANNENBAUM A., KIKINIS R.: On the laplace-beltrami operator and brain surface flattening. *IEEE Transactions on Medical Imaging 18*, 8 (1999), 700–711. 4

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (Nov. 2011), pp. 1626–1633. 4

[BHN03] BERMAN H., HENRICK K., NAKAMURA H.: Announcing the worldwide protein data bank. *Nature Structural & Molecular Biology 10*, 12 (12 2003), 980–980. 1

[BHW*19] BENHABILES H., HAMMOUDI K., WINDAL F., MELKEMI

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.

M., CABANI A.: A transfer learning exploited for indexing protein structures from 3d point clouds. In *Processing and Analysis of Biomedical Information*. Springer International Publishing, 2019, pp. 82–89. 3

[BMM*15] BOSCAINI D., MASCI J., MELZI S., BRONSTEIN M. M., CASTELLANI U., VANDERGHEYNST P.: Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. In *Computer Graphics Forum* (2015), vol. 34, Wiley Online Library, pp. 13–23. 4

[BSJ*01] BAKER N. A., SEPT D., JOSEPH S., HOLST M. J., MCCAMMON J. A.: Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences 98*, 18 (08 2001), 10037–10041. 2

[Can99] CANTERAKIS N.: 3d zernike moments and zernike affine invariants for 3d image analysis and recognition. In *In 11th Scandinavian Conf. on Image Analysis* (1999), pp. 85–93. 2

[CLM17] CRACIUN D., LEVIEUX G., MONTES M.: Shape Similarity System driven by Digital Elevation Models for Non-rigid Shape Retrieval. In *Eurographics Workshop on 3D Object Retrieval* (2017), Pratikakis I., Dupont F., Ovsjanikov M., (Eds.), The Eurographics Association. 4

[Con20] CONSORTIUM T. U.: UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research 49*, D1 (11 2020), D480–D489. 2

[ERXH*15] ESQUIVEL-RODRÍGUEZ J., XIONG Y., HAN X., GUANG S., CHRISTOFFER C., KIHARA D.: Navigating 3d electron microscopy maps with EM-SURFER. *BMC Bioinformatics 16*, 1 (05 2015). 3

[FBC13] FOX N. K., BRENNER S. E., CHANDONIA J.-M.: SCOPe: Structural classification of proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research 42*, D1 (12 2013), D304–D309. 2, 6

[GSM*19] GAINZA P., SVERRISSON F., MONTI F., RODOLÀ E., BOSCAINI D., BRONSTEIN M. M., CORREIA B.: Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods 17, pages184–192(2020)* (2019), 184–192. 1

[Kar11] KARPLUS M.: Behind the folding funnel diagram. *Nature Chemical Biology 7*, 7 (06 2011), 401–404. 1

[LPL*20] LANGENFELD F., PENG Y., LAI Y.-K., ROSIN P. L., ADERINWALE T., TERASHI G., CHRISTOFFER C., KIHARA D., BENHABILES H., HAMMOUDI K., CABANI A., WINDAL F., MELKEMI M., GIACHETTI A., MYLONAS S., AXENOPOULOS A., DARAS P., OTU E., ZWIGGELAAR R., HUNTER D., LIU Y., MONTÈS M.: SHREC 2020: Multi-domain protein shape retrieval challenge. *Computers & Graphics 91* (10 2020), 189–198. 2

[LW15] LIMBERGER F. A., WILSON R. C.: Feature encoding of spectral signatures for 3d non-rigid shape retrieval. In *BMVC* (2015), pp. 56–1. 4

[MCW*20] MISTRY J., CHUGURANSKY S., WILLIAMS L., QURESHI M., SALAZAR G., SONNHAMMER E. L. L., TOSATTO S. C. E., PALADIN L., RAJ S., RICHARDSON L. J., FINN R. D., BATEMAN A.: Pfam: The protein families database in 2021. *Nucleic Acids Research 49*, D1 (10 2020), D412–D419. 1, 6

[OZHL19] OTU E., ZWIGGELAAR R., HUNTER D., LIU Y.: Nonrigid 3d shape retrieval with happs: A novel hybrid augmented point pair signature. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (2019), pp. 662–668. 3

[QSMG16] QI C. R., SU H., MO K., GUIBAS L. J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593* (2016). 3

[RFB*21] RAFFO A., FUGACCI U., BIASOTTI S., ROCCHIA W., LIU Y., OTU E., ZWIGGELAAR R., HUNTER D., ZACHARAKI E. I., PSATHA E., LASKOS D., ARVANITIS G., MOUSTAKAS K., ADERINWALE T., CHRISTOFFER C., SHIN W.-H., KIHARA D., GIACHETTI A., NGUYEN H.-N., NGUYEN T.-D., NGUYEN-TRUONG V.-T., LE-THANH D., NGUYEN H.-D., TRAN M.-T.: Shrec 2021 track: Retrieval and classification of protein surfaces equipped with physical and chemical properties. *Computers & Graphics 99* (2021), 1–21. 4, 7

[RRBW*14] RODOLA E., ROTA BULO S., WINDHEUSER T., VESTNER M., CREMERS D.: Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 4177–4184. 4

[SLL*08a] SAEL L., LA D., LI B., RUSTAMOV R., KIHARA D.: Rapid comparison of properties on protein surface. *Proteins: Structure, Function, and Bioinformatics 73*, 1 (July 2008), 1–10. URL: https://doi.org/10.1002/prot.22141, doi:10.1002/prot.22141. 7

[SLL*08b] SAEL L., LI B., LA D., FANG Y., RAMANI K., RUSTAMOV R., KIHARA D.: Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins: Structure, Function, and Bioinformatics 72*, 4 (03 2008), 1259–1273. 1, 3

[SPNW04] SHULMAN-PELEG A., NUSSINOV R., WOLFSON H. J.: Recognition of functional sites in protein structures. *Journal of Molecular Biology 339*, 3 (06 2004), 607–633. 1

[TBG*20] THOMPSON E. M., BIASOTTI S., GIACHETTI A., TORTORICI C., AOUFEL WERGHI N., OBEID A. S., BERRETTI S., NGUYEN-DINH H.-P., A ND HAI-DANG NGUYEN M.-Q. L., TRAN M.-T., GIGLI L., VELASCO-FORERO S., TEGUI B. M., SIPIRAN I., BUSTOS B., ROMANELIS I., FOTIS V., TIS G. A., MOUSTAKAS K., OTU E., ZWIGGELAAR R., HUNTER D., LIU Y., ARTEAGA Y., LUXMAN R.: SHREC 2020: Retrieval of digital surfaces with similar geometric reliefs. *Computers & Graphics 91* (10 2020), 199–218. 4

[WSL*19] WANG Y., SUN Y., LIU Z., SARMA S. E., BRONSTEIN M. M., SOLOMON J. M.: Dynamic graph cnn for learning on point clouds. *ACM Transactions On Graphics 38*, 5 (2019), 1–12. 4, 5

[XZ09] XU D., ZHANG Y.: Generating triangulated macromolecular surfaces by euclidean distance transform. *PLoS ONE 4*, 12 (12 2009), e8140. 2

[Yuk15] YUKSEL C.: Sample elimination for generating poisson disk sample sets. *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2015) 34*, 2 (2015), 25–32. 3

[Zha05] ZHANG Y.: TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research 33*, 7 (04 2005), 2302–2309. 1

[ZLL*18] ZENG H., LIU Y., LI S., CHE J., WANG X.: Convolutional neural network based multi-feature fusion for non-rigid 3d model retrieval. *Journal of Information Processing Systems 14*, 1 (2018), 176–190. 4

[ZS04] ZHANG Y., SKOLNICK J.: Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics 57*, 4 (2004), 702–710. 1

[ZSSZ20] ZHANG Y., SUI X., STAGG S., ZHANG J.: FTIP: an accurate and efficient method for global protein surface comparison. *Bioinformatics 36*, 10 (02 2020), 3056–3063. 1

© 2021 The Author(s)
Eurographics Proceedings © 2021 The Eurographics Association.