Final Project - Comparing Mammalian DTI Scans with Unsupervised Optical Flow

Hila Cohen 31626460 [1]

[1] *Tel-Aviv University, Electrical Engineering, GDL Lab 2024*
hili4224@gmail.com
Guided by Mr. Gal Lifshitz & Dr. Dan Raviv

 Abstract— Deep learning models have emerged as a powerful tool for analyzing medical images, including tasks like disease classification and anomaly detection. In this paper, we explore their application to a different domain: comparing the brain structures of animals. Our study began with a 3D optical flow model initially developed for consecutive scans. We successfully adapted this model to handle multi-channel Diffusion Tensor Images (DTIs) of different animals. These modifications highlight the model's adaptability to higher-dimensional data and its ability to establish correspondences between points in distinct subjects. By leveraging the flow-derived points, we suggest a distance method to quantify structural differences between pairs of animals. This approach offers a quantitative framework for understanding brain variations across species.

 Keywords— Optical Flow, Unsupervised, Diffusion Tensor Image, Normalized Cross Correlation, Gromov-Hausdorff

I. INTRODUCTION

Optical flow estimation is a fundamental task in computer vision that aims to determine the apparent motion of pixels or features between consecutive frames or in a video sequence. This information plays a crucial role in various applications, such as object tracking [1] and action recognition [2]. Traditionally, optical flow has been estimated using classic approaches like Horn-Schunck [3] and Lucas-Kanade [4]. However, recent advancements in deep learning have led to the development of more sophisticated solutions There are several challenges that can make optical flow estimation a complex problem. These may include occlusions, large displacements, non-rigid motion and varying illumination [5].

A straightforward approach to tackling this task involves training a model on a dataset with known ground truth values. The FlowNet model [6] demonstrates the efficacy of convolutional neural networks (CNNs) in solving optical flow estimation as a supervised learning task.

DeepFlow [7], specifically designed to address the issue of large displacements, proposes a tailored descriptor matching algorithm that enhances performance in scenarios with fast motion. For occlusions, the ContinualFlow method [8] suggests estimating occlusions before flow, preventing contamination of the final output by flow corrupted as a result of occlusions.

However, the supervised approach can be hindered by the limited availability of real-world datasets with ground truth annotations. This has prompted researchers to explore unsupervised learning methods.Guided Optical Flow Learning [9] proposes a novel framework that employs proxy ground truth data generated from classical approaches to guide CNN learning. The Back to Basics approach [10] leverages a loss function that combines a data term (measuring photometric constancy over time) with a spatial term (modeling the expected variation of flow across the image.

This work explores the application of optical flow estimation to diffusion tensor imaging (DTI) scans of different animals. Our aim is to identify corresponding points and estimate the similarity between these scans. We build upon the Final Project of Yariv Levy and Doron Kopit, from the GDL lab, which focused on cardiac volumetric optical flow estimation using DNN. Their model architecture is based on the Light version of PWC-Net [11], as presented in the ARFlow paper [12]. The proposed distance calculation for comparing the scans is inspired by the Gromov-Hausdorff metric [13].

Several studies have explored medical applications of optical flow and DTI registration. The FlowReg model [14] combines two neural networks: FlowReg-A for affine transformation and FlowReg-O for 2D optical flow-based refinement. VoxelMorph [15] offers an unsupervised framework for medical data registration using cross-correlation loss.

For DTI-specific registration, The paper [16] proposed a registration-based approach for up sampling DTI images. The work in [17] presented a method that incorporates explicit tensor reorientation optimization. This approach leverages conjugate gradients for the optimization problem.

While most existing studies focus on optical flow between objects of the same type or within the same scene, our project investigates the application of optical flow to DTI scans of entirely different mammals. We adapted a pre-existing optical flow model, originally designed for more traditional tasks as stated previously, to this unique challenge. Through experimentation, we identified a configuration for achieving visually satisfactory image warping results.
Git link can be found here [1]

II. DATA

Diffusion tensor imaging is an advanced magnetic resonance imaging modality that uses the Brownian motion of water molecules to provide data for images [18]. The data generated by DTI characterizes water diffusion using an effec-

---

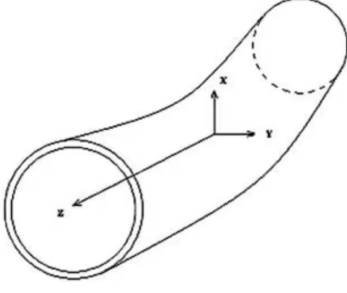[1] https://github.com/Hila954/Project_DTI

Fig. 1: Diffusion ellipsoid. Three eigenvectors are demonstrated, with the principal eigenvector along the Z direction, origin [22].

tive diffusion tensor, D [19]. This tensor allows for the direct examination of brain microstructure, making DTI a valuable tool for studying various neurological disorders, including stroke, epilepsy, multiple sclerosis, brain tumors, and Alzheimer's [20]. For each estimated D, whether at the tissue or voxel level, a local orthogonal coordinate system (principal coordinate axes) can be constructed. Additionally, three corresponding diffusion coefficients (principal diffusivities) can be calculated along these axes, representing the eigenvalues of the tensor. The eigenvector associated with the largest eigenvalue defines the tissue's fiber-tract orientation, while the other two eigenvectors represent the remaining orthotropic axes [21]. An example is illustrated in Figure 1. Because D is symmetric and positive definite, the DTI data of each pixel in each scan contains 6 elements, related to the estimated matrix. Thus, we get a **3D image with 6 channels**.

III. THE DISTANCE

Given a pair of DTI images, one for each animal, our goal is to compute a similarity value. The proposed approach is inspired by the Gromov-Hausdorff distance [13]. We will first give the formal definition and then the approximation.

Definition of Gromov-Hausdorff. *Given $(X, dX), (Y, dY) \in M$, where M denotes the set of all compact metric spaces, the GH distance measures how far the two metric spaces are from being isometric. It considers any "sufficiently rich" third metric space (Z, dZ) that contains isometric copies of X and Y, measuring the Hausdorff distance (in Z) between these copies, and minimizes over the choice of the isometric copies and Z [23].*

$$d_{GH}(X,Y) = \inf_{Z, \varphi_X, \varphi_Y} d_Z^H(\varphi_X(X), \varphi_Y(Y)) \qquad (1)$$

where φ_X and φ_Y are isometric embeddings of X and Y into Z
d_Z^H is the Hausdorff distance in Z:

$$d_Z^H(S,T) = \max\left\{\sup_{s \in S}\inf_{t \in T} d_Z(s,t), \sup_{t \in T}\inf_{s \in S} d_Z(s,t)\right\} \quad \forall S, T \subseteq Z. \qquad (2)$$

The initial approach to comparing metric spaces involves finding a common metric space in which both spaces can be viewed as subspaces. In this shared space, their similarity can be measured using the Hausdorff distance [24]. The Gromov-Hausdorff (GH) distance offers a natural measure of similarity in the context of shape recognition. As such, it is a promising candidate for defining shape distances [25].

The notion of isometry can be relaxed in order to define similarity of shapes. We will refer to a set $C \subseteq X \times Y$ of pairs such that for every $x \in X$ there exists at least one $y \in Y$ such that $(x, y) \in C$, and similarly for every $y \in Y$ there exists an $x \in X$ such that $(x, y) \in C$, as a correspondence between X and Y.

The GH distance can also be described in terms of the distortion resulting from embedding one surface into another [26]:

$$d_{GH}(X,Y) := \frac{1}{2}\inf_C \operatorname{dis}(C) \qquad (3)$$

where the infimum is taken over all correspondence C.
dis(C) is the distortion defined:

$$\operatorname{dis}(C) := \sup_{(x,y),(x',y') \in C} |d_X(x,x') - d_Y(y,y')| \qquad (4)$$

Once we have identified corresponding points using the calculated flow, we can proceed to calculate the aforementioned approximation. Within the DTI image, the distance between points is defined as the shortest path connecting the source and target points. To incorporate the influence of intensity and location differences, we model the image as a weighted graph. The weight associated with an edge between neighboring vertices is determined by the following formula:

$$dl = \sqrt{dx^2 + dy^2 + dz^2 + \lambda\left(\sum_{i=1}^{6} dI_i^2\right)} \qquad (5)$$

Where $[dx, dy, dz]$ are the voxel dimensions and dI_i is the matching image gradient for channel i.
The final distance is calculated using the following process:

Step 1 - Select a random voxel in image A as the anchor point p, and define a window of a specified size around it

Step 2 - Calculate the distance between the anchor point p and each point within the window [x1...xn]. We employ the Dijkstra algorithm [27] to efficiently determine the shortest paths between nodes in the weighted graph

Step 3 - Identify the corresponding anchor point p' in image B and calculate the distance from p' to each of the matching voxels [x1'...xn']. The voxel xi' in image B corresponds to the voxel xi in image A.

We denote the output of step 2 as matrix D1 and the output of step 3 as matrix D2, as illustrated in Figure 2 We substitute both matrices into the formula 6 to calculate the current distance.

$$Distance = \sqrt{E[(D1 - D2)^2]} \qquad (6)$$

To enhance noise resilience, we approximate the supremum and infimum operations with average. To obtain the final result, we repeat this process for n anchor points across image A and calculate the average.

IV. THE OPTICAL FLOW MODEL

Building upon the Final Project of Yariv Levy and Doron Kopit, we utilize the same model architecture and similar configurations. The network is based on the lightweight multi-frame extension of PWC-Net [11], as proposed in [12]. We will subsequently explore the neural network architecture and loss functions.

distances in image1 – D1

	X1	X2	X3	.....
ANCHOR POINT P	MIN_DISTANCE FROM ANCHOR POINT TO X1	MIN_DISTANCE FROM ANCHOR POINT TO X2		

distances in image2 – D2

	X1'	X2'	X3'	.....
ANCHOR POINT P'	MIN_DISTANCE FROM ANCHOR POINT TO X1'	MIN_DISTANCE FROM ANCHOR POINT TO X2'		

Fig. 2: Each matrix represents the distances from the anchor point to corresponding points in the respective images

a. Architecture

The model architecture is illustrated in 3. It consists of several components.

The Pyramiding Feature Extractor The Pyramiding Feature Extractor generates feature representations at each layer L. Convolutional filters are applied to down sample the image from the previous layer (L-1) to produce a total of 7 feature maps, with varying number of channels, for each image. The original input image serves as the bottom level. Unlike the original model, which used a 3D image with 3 channels, we now have 6 channels to accommodate DTI data

The next main steps are performed in a loop over the pyramid feature maps:

The Warping Layer Warp the current feature map of image 2 using the up sampled flow of the previous map, towards the current feature map of image 1. If no flow has been calculated yet, the current map remains unchanged.

The Cost Volume Layer correlation operation between the wrapped feature map of image 2 and the current feature map of image 1.

The Optical Flow Estimator a multi-layer CNN, the input is organized by concatenating the feature maps of source image 1 , the up sampled flow from the previous higher level and the cost volume.

The Context Network employed to refine the optical flow by leveraging contextual information. This network is a feed-forward CNN composed of 7 convolutional layers and is designed using dilated convolutions.

By the end of the pyramid iteration, the network predicts optical flow at different scales. in practice, the loop ends after 5 levels and we up sample the flow by 4.
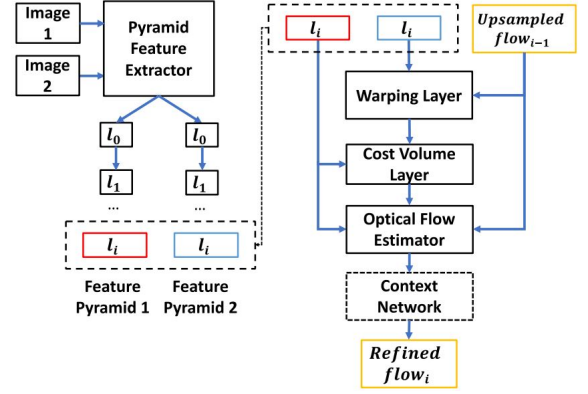

Fig. 3: The Network architecture, a light version of PWC-Net, taken from Yariv & Doron's Final Project report

b. Training Aspects

In this section we will review loss functions and training methods that were employed through the process. It is important to note that the training is being done for both flow directions, image A to B and image B to A.

1. Normalized Cross Correlation

Correlation is a widely used similarity measure in matching tasks. Compared to the classic Mean Square Error, cross-correlation is more robust to intensity variation found across scans and datasets [28]. This loss function serves as the primary training objective for our model, as demonstrated in the experimental results. The NCC formula can be found in 7. This loss is also refered as Photometric Loss

$$NCC = \left(\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left\{ A_{jk} - \bar{a} \right\} \left\{ B_{jk} - \bar{b} \right\}}{\sigma_A * \sigma_B} \right)^2 \tag{7}$$

Where K and J are the window dimensions around a pixel, $\bar{a}$ ($\bar{b}$) is the average of image A(B) in this window. σ_A (σ_B) denotes the corresponding standard deviation. As we deal with multi-channel problem, in practice we calculate the average NCC across the channels, over the image.

2. Smoothness Loss

A common regularization technique in optical flow estimation is the smoothness constraint. This constraint assumes that neighboring pixels belonging to the same object exhibit consistent motion, resulting in a smooth flow field. While our application doesn't involve rigid object motion, we found that incorporating a smoothness constraint into the training loss significantly improves the results. The smoothness regulation is shown in 8, as stated in [12].

$$L_{sm} \sim \sum_{d \in \{x,y,z\}} \sum_{p} \|\nabla_d U_{12}\|_1 \, e^{-|\nabla_d I|} \tag{8}$$

U_{12} represents the optical flow and I the origin image. This constrains the flow to be smooth (similar) in x, y and z when no significant image gradient exists. In our case, we do not consider the intensity gradients and set the exponent to 1

3. Cycle Loss

By aiming to estimate the flow from one image to another and vice versa, we can leverage cycle consistency as a form of self-supervision for the mode, as set in [29]. Formula shown in 9

$$L^C = \sum_{\mathbf{x} \in N} \left(\mathbf{F}_{i \to j}(\mathbf{x}) + \mathbf{F}_{j \to i}(\mathbf{x}) \right)^2 \qquad (9)$$

As $\mathbf{F}_{i \to j}$ is the flow prediction, mapping scan I_i to scan I_j, and N is the set of non-occluded voxel locations.

4. Occlusion Mask

Occlusions and out-of-view pixels can hinder the training process by preventing the establishment of corresponding points. To address this, the original ARFlow paper proposed an occlusion map. As described in [12], this map is generated using the classical forward-backward checking method, where the backward flow is estimated by reversing the order of input images. The NCC loss is then discarded for occluded regions by pre-multiplying the input images with the generated mask.

V. EXPERIMENTS

This work focuses on 10 DTI scans from various animals, each representing a 3D space with 6 channels. All data is reshaped to [192, 192, 128] using zero padding and rescaled to the range of [0,1]. To account for inter-species differences, we include the voxel dimensions of each scan alongside the DTI values as part of our dataset.

It's essential to note that our goal is not to create a universal model applicable to all animal pairs. Instead, we focus on using optical flow to identify corresponding points for a specific pair and subsequently apply the proposed distance calculation. Real-time performance is not a priority for this task.

a. Feasibility check – toy problem

Pretraining – Changing window size   To evaluate the model's applicability to 6-channel images, we began with a straightforward experiment. We used DTI data of a Hyrax and shifted it by 5 pixels along the z-axis. The original Hyrax and its shifted version were then combined as 2 frames input to the model. We train the model from scratch in an overfitting manner.

A simple training configuration was employed, using only the NCC loss for 3000 iterations. For the initial training run, we used a dynamic window size for NCC loss calculation, increasing the window by 2 after every 1000 iterations.

To evaluate performance, we applied the optical flow output to warp the shifted image back to its original state and vice versa, result in Fig.4. The model accurately captured the shift and successfully reconstructed each image from the other.

For this specific case, we could also validate performance by comparing against the ground truth. Knowing the exact pixel shift allowed us to determine the expected optical flow. We used Mean Squared Error (MSE) for evaluation, with results shown in Figure 5. It's important to note that occluded pixels were excluded from the MSE calculation to avoid their impact on the metric. Despite visually acceptable results,
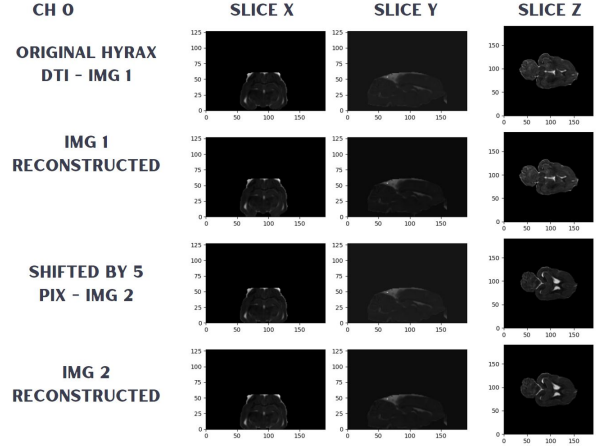

Fig. 4: We present the 3D image using the middle slice in each axis, the data belongs to channel 0. The model managed to reconstruct each image from the other
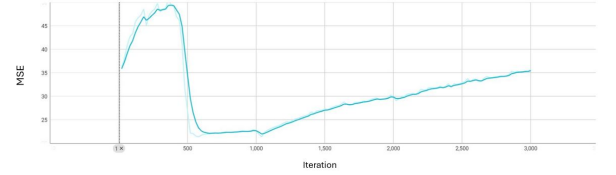

Fig. 5: MSE for toy example, pretrain stage

the MSE indicates that the model is struggling with occluded patches, as evidenced by the increasing error value.

Finetune   Starting with the pretrained model from the previous step, we introduced the occlusion mask, cycle loss, and smoothness loss 10 for an additional 4000 iterations. Figure 6 demonstrates the new wrapped results and 7 the corresponding MSE. As anticipated, these modifications enabled the model to optimize the flow and achieve a much closer approximation of the ground truth.

$$L_{Finetune} = L_{NCC} + L_{sm} + L_{cyc}(+occlusion) \qquad (10)$$

we add cycle loss for all flow levels except the smallest, and smoothness to the last level flow only.
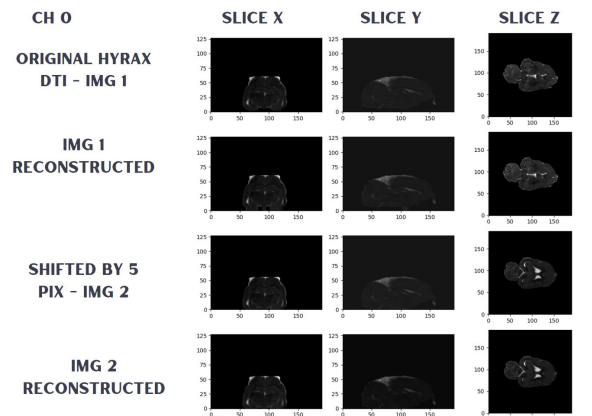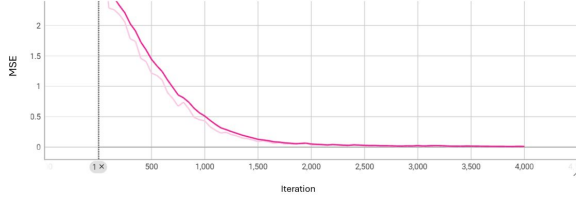

Fig. 6: Wrapping using the refined flow.
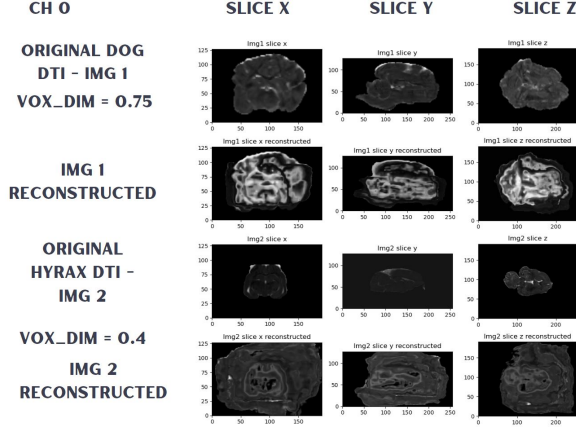
Fig. 7: Improved MSE after finetune

Fig. 8: Dog vs Hyrax Results after resample - Pretrain. The reconstructed Dog scan appears overly sparse, while the reconstructed Hyrax scan contains numerous extraneous elements.

b. Real Data – Hyrax vs Dog

Resample Approach

Pretraining Having demonstrated the model's effectiveness for 6-channel images, we proceeded to evaluate its performance on a real pair of DTI scans: a Dog and a Hyrax. This stage differs from typical optical flow estimation in that it involves non-consecutive frames representing two distinct brains. Our goal remains to accurately estimate the optical flow, which can be used to warp one image onto the other for visual evaluation.

Given the differing voxel dimensions between the hyrax (0.4mm) and dog (0.75mm) scans, we resampled the data to a common voxel dimension of 0.4mm to ensure comparability. The model was trained from scratch using only the NCC loss with a window size of 3, 6000 iterations. The results are presented in Figure 8. Due to the size differences between the brains, many voxels remain unmatched, indicating a lack of corresponding points. For this resampling experiment, the padded size was set to [256, 192, 128]

Finetune To refine the flow, we incorporated cycle loss, smoothness, and the occlusion map, as done in the toy example. However, during training, we observed that the flow struggled to maintain cycle consistency, leading to an increasing number of pixels being classified as occluded. This resulted in a significant portion of the image being excluded from the loss calculation, effectively hindering the model's training. The flawed flow output is illustrated in Figure 9.
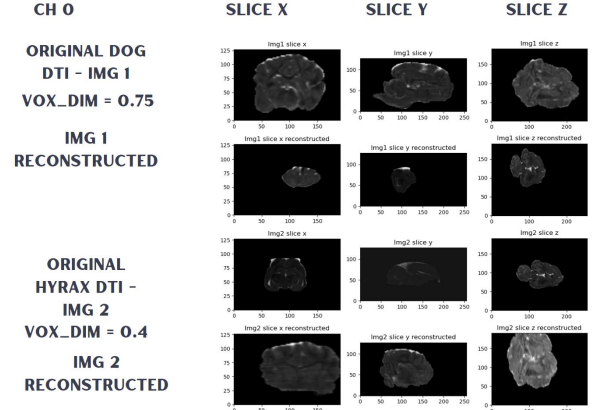

Fig. 9: Dog vs Hyrax Results after resample - Finetune. The occlusion mask harmed the training process and the output is flawed.
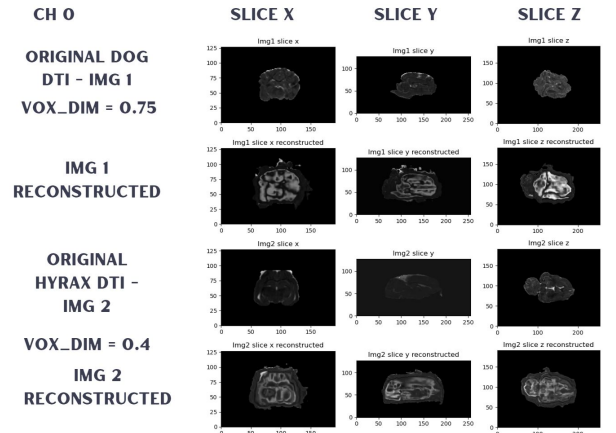

Fig. 10: Dog vs Hyrax Results no resample - Pretrain.

No-Resample

Pretrain In this experiment, we maintained the original voxel dimensions of both scans without attempting to match them. The model was trained from scratch using only the NCC loss with a window size of 3, 6000 iterations. The results are depicted in Figure 10.

We observe a reduction in unmatched pixels compared to the previous experiment, indicating improved performance. However, it's important to note that disregarding the physical dimensions of the brain during optical flow estimation may compromise real-world accuracy. To mitigate this potential limitation, we plan to compensate for it in the subsequent distance calculation stage.

Finetune Starting with the pretrained model, we introduce a smoothness loss of 0.1 for all flow levels while omitting cycle loss and the occlusion mask. The results are illustrated in Figure 11. The smoothness loss effectively refined the flow, leading to a significant improvement in image reconstruction.

$$L_{Finetune} = L_{NCC} + L_{sm} \qquad (11)$$

We trained for another 6000 iterations, NCC window size 3. The same finetuned process can be performed for different pairs of animals, examples in figures 12 and 13
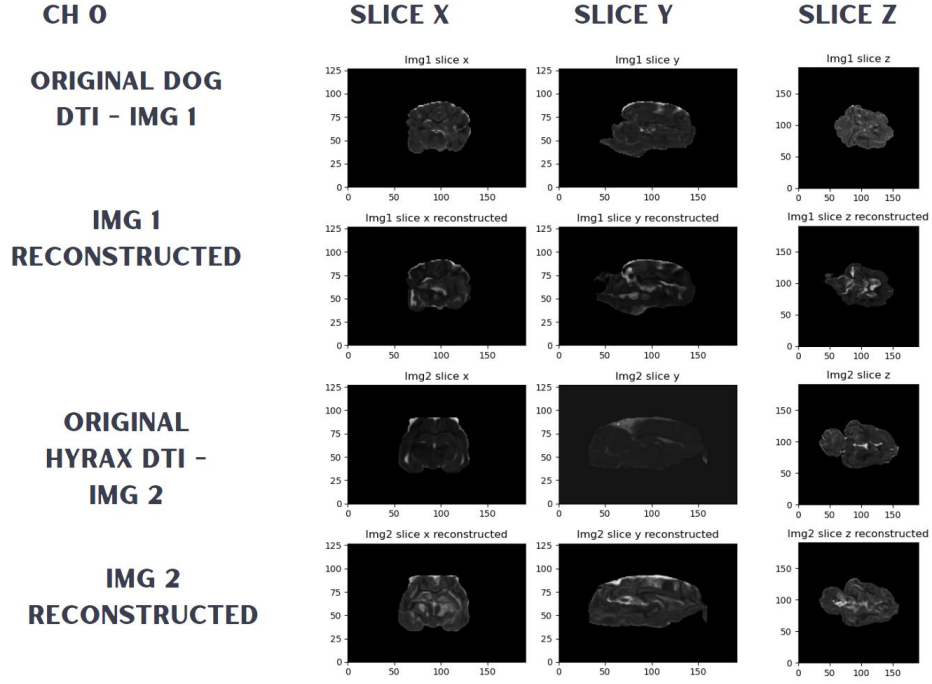
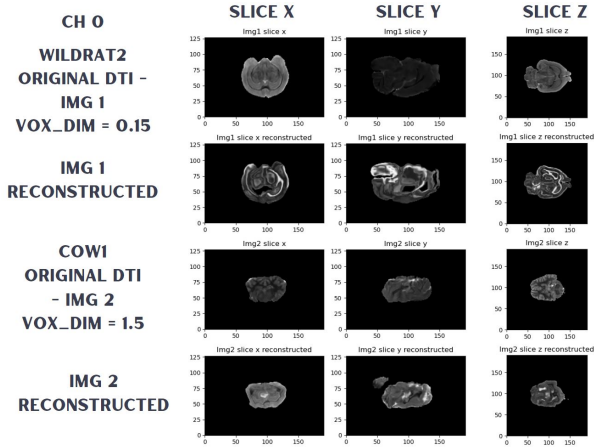**Fig. 11:** Dog vs Hyrax results no resample - Finetune.

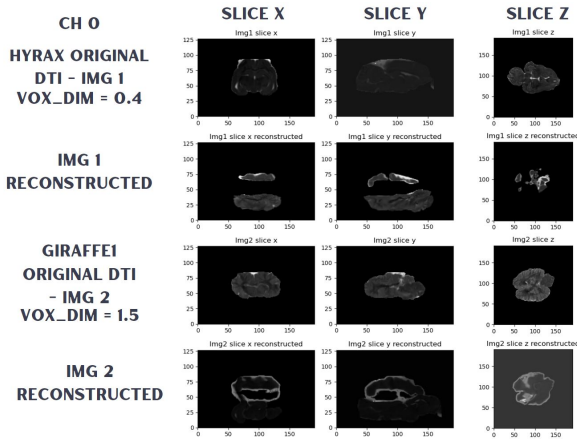**Fig. 12:** WildRat vs Cow results no resample - Finetune.

Fig. 13: Hyrax vs Giraffe results no resample - Finetune. In this example, the model failed to predict the desired flow, resulting in a poor output. Many voxels in the z-slice (for img 1 reconstruction) were incorrectly warped beyond the image boundaries.

VI. RESULTS - COMPUTING THE DISTANCES

In the final step, we fine-tune the model for each pair of animals within our dataset. By combining the 10 available animal options, we can explore various pairings. We fine tuned most pairs for about 1500 iterations, using the configuration explained in b, the initial weights are the output of this section. The resulting optical flow allows us to identify corresponding points between each pair and then calculate the Gromov-Hausdorff distance. In the absence of ground truth or a specific validation method, we evaluated our results based on biological knowledge. Our dataset includes three pairs of animals that are expected to exhibit a closer relationship than the others:

- **Dog vs Wolf** - The dog is a subspecies of the gray wolf [30]

- **Orangutan vs Chimpanzee** - Orangutans and chimpanzees are both apes [31] [32]

- **Donkey vs Horse** - All donkeys and horses evolved from a common ancestor, Dinohippus [33]

We anticipate that the results will align with the expected relationships between these pairs, demonstrating their greater similarity compared to other combinations.

As detailed in the III, this process involves several adjustable parameters. The key parameters are:

- **The window size** Determining the optimal window size around the anchor point is a crucial consideration. A larger window can improve the distance accuracy, but it also increases computational cost.

- **The λ value** The chosen parameter directly impacts the intensity penalty applied to dijkstra paths, it determines how strict we are gradient-wise when searching for the shortest path.

- **Number of anchor points** To ensure that the estimated distance is independent of the chosen anchor point, and to better generalize the method, we calculate the average distance across multiple anchors

Figure 14 illustrates the results of averaging across 50 anchor points, using a window size of 5 pixels in each dimension and a lambda value of 0. For this configuration, we excluded the impact of intensity.

Our results demonstrate that even without considering image gradients, the Euclidean-based method effectively distinguishes the expected pairs from the others, confirming their closer relationship. However, the Wolf-Dog pair exhibited an unexpected result, The Wolf scan is closer to the Hyrax instead of the Dog. Upon examining the raw DTI data for the Wolf, we identified that there may be a flaw in the scan. In addition, because of this flaw - the model struggled when the Wolf scan is used.

Another type of possible visualization, showed in 15, is the use of Multidimensional Scaling, that aims to realize a given matrix of dissimilarities, as n points in a k-dimensional Euclidean space. [34]

Figure 16 presents the results obtained by averaging across 50 anchor points, using a window size of 9 pixels in each dimension and a λ value of 200. The calculated distances between the known similar pairs remain significantly smaller than those of other pairs. Furthermore, the gap between the smallest distances and the rest has widened. This positive outcome likely stems from the larger window size, which incorporates information from more voxels.

VII. CONCLUSIONS

This study successfully applied an optical flow neural network to estimate distances between DTI scans of various mammals. We demonstrated the model's ability to handle 6-channel inputs and address the challenges of estimating optical flow between distinct subjects.

The proposed distance approach, inspired by the Gromov-Hausdorff metric, effectively captured the expected similarities between designated pairs within our dataset, with the exception of the potentially problematic Wolf DTI scan.

Future research directions may include:

- Expanding the dataset to include more animals and optimizing the distance parameters.

- Developing a quantitative validation metric for the task of warping one DTI image onto another, as our current evaluation relies primarily on visual inspection.

- Investigating alternative model parameters to address the unsatisfactory warping results observed in the Hyrax involved pairings.

REFERENCES

[1] Q. Wang, Y.-Y. Chang, R. Cai, Z. Li, B. Hariharan, A. Holynski, and N. Snavely, "Tracking everything everywhere all at once," in *International Conference on Computer Vision*, 2023.

[2] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *CoRR*, vol. abs/1406.2199, 2014. [Online]. Available: http://arxiv.org/abs/1406.2199

[3] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0004370281900242

[4] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision (ijcai)," vol. 81, 04 1981.

[5] S. Shah and X. Xuezhi, "Traditional and modern strategies for optical flow: an investigation," *SN Applied Sciences*, vol. 3, 03 2021.

[6] P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," *CoRR*, vol. abs/1504.06852, 2015. [Online]. Available: http://arxiv.org/abs/1504.06852

[7] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.

[8] M. Neoral, J. Sochman, and J. Matas, "Continual occlusions and optical flow estimation," *CoRR*, vol. abs/1811.01602, 2018. [Online]. Available: http://arxiv.org/abs/1811.01602

[9] Y. Zhu, Z. Lan, S. D. Newsam, and A. G. Hauptmann, "Guided optical flow learning," *CoRR*, vol. abs/1702.02295, 2017. [Online]. Available: http://arxiv.org/abs/1702.02295

[10] J. J. Yu, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *Computer Vision – ECCV 2016 Workshops*, G. Hua and H. Jégou, Eds. Cham: Springer International Publishing, 2016, pp. 3–10.

[11] D. Sun, X. Yang, M. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jun 2018, pp. 8934–8943. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00931

[12] L. Liu, J. Zhang, R. He, Y. Liu, Y. Wang, Y. Tai, D. Luo, C. Wang, J. Li, and F. Huang, "Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation," *CoRR*, vol. abs/2003.13045, 2020. [Online]. Available: https://arxiv.org/abs/2003.13045

[13] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, 12 2006, vol. 152, pp. –586.

[14] S. Mocanu, A. R. Moody, and A. Khademi, "Flowreg: Fast deformable unsupervised medical image registration using optical flow," *CoRR*, vol. abs/2101.09639, 2021. [Online]. Available: https://arxiv.org/abs/2101.09639

[15] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. V. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *CoRR*, vol. abs/1809.05231, 2018. [Online]. Available: http://arxiv.org/abs/1809.05231

[16] Z. Mai, M. Verhoye, A. V. d. Linden, and J. Sijbers, "Diffusion tensor images upsampling: A registration-based approach," in *2009 13th International Machine Vision and Image Processing Conference*, 2009, pp. 36–40.

[17] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor mr images with explicit orientation optimization," *Medical Image Analysis*, vol. 10, no. 5, pp. 764–785, 2006, the Eighth International Conference on Medical Imaging and Computer Assisted Intervention – MICCAI 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841506000417

[18] S. T. Ranzenberger LR, Das JM, "Diffusion tensor imaging," [Updated 2023 Nov 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537361/, accessed: 2024-09-07.

[19] D. of Medical Physics University of Wisconsin School of Medicine and P. Health, "Diffusion tensor imaging (dti)," https://medphysics.wisc.edu/research/fmri/diffusion-tensor-imaging/, accessed: 2024-09-07.

[20] J. Soares, P. Marques, V. Alves, and N. Sousa, "A hitchhiker's guide to diffusion tensor imaging," *Frontiers in Neuroscience*, vol. 7, p. 31, 03 2013.

[21] P. J. Basser, J. Mattiello, and D. LeBihan, "Mr diffusion tensor spectroscopy and imaging." *Biophysical journal*, vol. 66 1, pp. 259–67, 1994. [Online]. Available: https://api.semanticscholar.org/CorpusID:8747275

	Dog	Hyrax	WildRat2	Cow1	Giraffe1	Orangutan1	Donkey	Chimpanzee	Horse1	Wolf1
Dog	0.0	1.23	3.06	5.33	2.95	4.07	4.51	4.57	2.83	1.52
Hyrax	1.23	0.0	1.88	4.32	4.02	5.79	3.48	5.27	3.95	1.32
WildRat2	3.06	1.88	0.0	6.47	7.08	5.73	6.47	5.18	6.89	1.87
Cow1	5.33	4.32	6.47	0.0	2.84	3.63	2.91	3.21	2.38	4.3
Giraffe1	2.95	4.02	7.08	2.84	0.0	2.82	3.37	3.2	2.82	4.49
Orangutan1	4.07	5.79	5.73	3.63	2.82	0.0	2.81	2.16	2.5	3.99
Donkey	4.51	3.48	6.47	2.91	3.37	2.81	0.0	3.02	2.26	4.4
Chimpanzee	4.57	5.27	5.18	3.21	3.2	2.16	3.02	0.0	2.73	4.58
Horse1	2.83	3.95	6.89	2.38	2.82	2.5	2.26	2.73	0.0	5.07
Wolf1	1.52	1.32	1.87	4.3	4.49	3.99	4.4	4.58	5.07	0.0

Fig. 14: This distance matrix was calculated using a window size of 5, averaging across 50 anchor points, and excluding the influence of gradients. As observed in previous analyses, the optical flow estimation was unreliable for certain pairings. To highlight these cases, we marked the extreme outliers with a red underline.
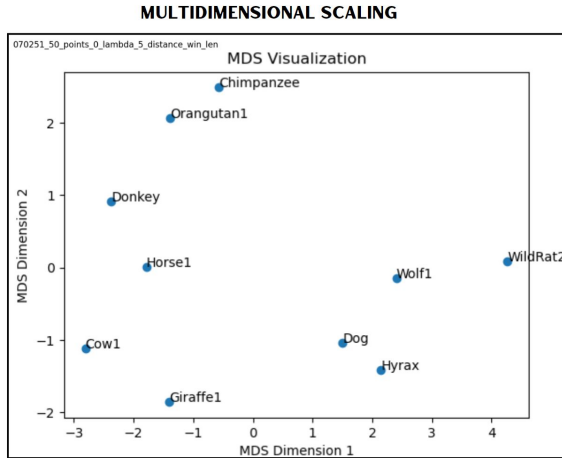
MULTIDIMENSIONAL SCALING

Fig. 15: 2-D MDS visualization for window of 5 and $\lambda = 0$. The MDS algorithm aims to find a vector representation for object, so the original distances will be preserved.

The Third International Workshop on Biomedical Image Registration – WBIR 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841507000606

[29] G. Lifshitz and D. Raviv, "Unsupervised volumetric displacement fields using cost function unrolling," in *Biomedical Image Registration, Domain Generalisation and Out-of-Distribution Analysis*, M. Aubreville, D. Zimmerer, and M. Heinrich, Eds. Cham: Springer International Publishing, 2022, pp. 153–160.

[30] C. B. Vanacore, "Dog," *Encyclopedia Britannica*, September 2024. [Online]. Available: https://www.britannica.com/animal/dog

[31] T. Nishida, "Chimpanzee," *Encyclopedia Britannica*, September 2024. [Online]. Available: https://www.britannica.com/animal/chimpanzee

[32] B. M. F. Galdikas, "Orangutan," *Encyclopedia Britannica*, September 2024. [Online]. Available: https://www.britannica.com/animal/orangutan

[33] S. L. o. t. B. Ben Hart and T. D. S. Human Behaviour department, "What's the difference between a horse and a donkey?" https://www.discoverwildlife.com/animal-facts/mammals/donkey-vs-horse-whats-the-difference, accessed: 2024-09-16.

[34] A. Boyarski and A. Bronstein, *Multidimensional Scaling*.

[22] D. U. C. o. M. Avneesh Chhabra MD Staff Radiologist, Department of Radiology, "Diffusion tensor imaging," https://emedicine.medscape.com/article/345561-overview?form=fpf, accessed: 2024-09-07.

[23] V. Oles, N. Lemons, and A. Panchenko, "Efficient estimation of a gromov–hausdorff distance between unweighted graphs," 11 2019.

[24] F. Schmiedl, "Computational aspects of the gromov—hausdorff distance and its application in non-rigid shape matching," *Discrete Comput. Geom.*, vol. 57, no. 4, p. 854–880, jun 2017. [Online]. Available: https://doi.org/10.1007/s00454-017-9889-4

[25] F. Mémoli and G. Sapiro, "A theoretical and computational framework for isometry invariant recognition of point cloud data," *Found. Comput. Math.*, vol. 5, no. 3, p. 313–347, jul 2005.

[26] A. Bronstein, M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro, "A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching," *International Journal of Computer Vision*, vol. 89, pp. 266–286, 09 2010.

[27] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, p. 269–271, dec 1959. [Online]. Available: https://doi.org/10.1007/BF01386390

[28] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008, special Issue on

092057_50_points_200_lambda_9_distance_win_len

	Dog	Hyrax	WildRat2	Cow1	Giraffe1	Orangutan1	Donkey	Chimpanzee	Horse1	Wolf1
Dog	0.0	2.21	5.8	8.57	4.98	6.67	7.59	7.27	4.82	3.07
Hyrax	2.21	0.0	4.11	7.27	7.07	10.21	5.97	9.41	7.07	2.41
WildRat2	5.8	4.11	0.0	11.12	12.17	9.74	11.28	8.67	12.2	3.07
Cow1	8.57	7.27	11.12	0.0	4.78	6.12	5.09	5.81	4.19	7.07
Giraffe1	4.98	7.07	12.17	4.78	0.0	4.58	6.09	5.2	4.83	7.3
Orangutan1	6.67	10.21	9.74	6.12	4.58	0.0	4.76	3.6	4.39	6.21
Donkey	7.59	5.97	11.28	5.09	6.09	4.76	0.0	4.76	3.87	7.55
Chimpanzee	7.27	9.41	8.67	5.81	5.2	3.6	4.76	0.0	4.56	7.1
Horse1	4.82	7.07	12.2	4.19	4.83	4.39	3.87	4.56	0.0	8.45
Wolf1	3.07	2.41	3.07	7.07	7.3	6.21	7.55	7.1	8.45	0.0

Fig. 16: This distance matrix was calculated using a window size of 9, averaging across 50 anchor points, and a λ of 200. The majority of distances increased, resulting in a wider gap between the pair with the smallest distance and the others.