Geometry-Aware Hemodynamics via a Transformer Encoder and Anisotropic RBF Decoder

Reza Akbarian Bafghi*
University of Colorado, Boulder
reza.akbarianbafghi@colorado.edu

Sukirt Thakur*
AngioInsight, Inc
sukirt@angioinsight.com

Maziar Raissi
University of California, Riverside
maziar.raissi@ucr.edu

Abstract

Accurate and rapid estimation of hemodynamic metrics, such as pressure and wall shear stress (WSS), is essential for diagnosing and managing Coronary Artery Disease (CAD). Existing approaches, including invasive Fractional Flow Reserve (FFR) measurements and computationally expensive Computational Fluid Dynamics (CFD) simulations, face challenges in invasiveness, cost, and speed. We present a framework that accelerates non-invasive coronary hemodynamics prediction. The model integrates 1D centerline and inlet flow rate into a transformer-based encoder, followed by an anisotropic Radial Basis Function (RBF) decoder that aligns with vessel morphology for continuous wall-based predictions. We also introduce a large synthetic dataset of 4,000 single-vessel coronary artery geometries with corresponding steady-state flow simulations, enabling robust training and evaluation. Our method improves accuracy and delivers orders-of-magnitude speedups over CFD on this synthetic benchmark. While tested only on steady, single-vessel cases, it shows promise for clinical acceleration; validation on clinical data and extension to multi-vessel and transient settings are important next steps. Dataset available at: https://huggingface.co/datasets/angioinsight/single-vessel-flow

1 Introduction

Coronary artery disease (CAD) remains a leading cause of morbidity and mortality worldwide [1], necessitating accurate, timely diagnostic tools. Fractional flow reserve (FFR), the functional gold standard [2], guides treatment by quantifying the hemodynamic impact of stenoses, but its invasive nature limits broad clinical use [3].

Computational Fluid Dynamics (CFD) offers a non-invasive alternative by simulating patient-specific blood flow to estimate FFR, pressure, and wall shear stress (WSS) [4]. However, CFD's high computational cost, often several CPU-hours to days per case [5], restricts real-time application. This challenge has spurred machine learning approaches to accelerate hemodynamic prediction [6].

Approaches include physics-informed neural networks (PINNs) [7, 8] and neural operators [9]. PINNs enforce PDE constraints directly but face challenges with nonlinear, multi-scale flows [10, 11]. Neural operators like Fourier Neural Operator (FNO) [12] and geometry-aware variants [13, 14] learn resolution-independent mappings. Transformer-based operators [15–18] improve flexibility but often lack explicit geometric and hemodynamic priors needed for vascular anatomies. Cardiovascular flow poses unique challenges: intricate vessel geometries, stenoses, multi-scale features, and patient-specific hemodynamics.

*These authors contributed equally.

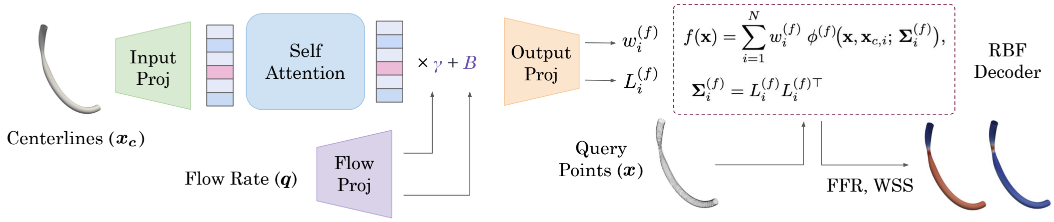

Figure 1: Architecture of the Transformer-Anisotropic RBF Network. The transformer encoder processes centerline geometry and flow rate, producing parameters for anisotropic RBF kernels aligned with vessel morphology to reconstruct continuous pressure fields.

Our work addresses these challenges by integrating 1D centerline and inlet flow rate into a transformer encoder, paired with an anisotropic Radial Basis Function (RBF) decoder aligned to local vessel morphology. The core contributions include:

- A transformer encoder with 4D Fourier embeddings of centerline geometry and radius, paired with an anisotropic RBF decoder for continuous wall predictions.
- A large-scale synthetic coronary hemodynamics dataset comprising 4,000 single-vessel coronary artery geometries with corresponding steady-state CFD simulations, enabling robust training, fair benchmarking, and reproducible evaluation. This open dataset provides a standardized testbed for developing and evaluating hemodynamic prediction models, including FFR and wall shear-based risk assessment.

2 Methodology

Our proposed Transformer-Anisotropic RBF Network predicts continuous pressure fields on coronary artery walls from vessel geometry and inlet flow rate. The framework combines a transformer-based encoder for geometry-aware feature extraction with an anisotropic Radial Basis Function (RBF) decoder for continuous field reconstruction (Fig. 1).

2.1 Problem Formulation

We aim to learn a mapping from vessel geometry, hemodynamic descriptors, and boundary conditions to wall pressure. The input is a vessel centerline $\mathbf{C} = \{(\mathbf{x}_i, r_i)\}_{i=1}^N$ with coordinates $\mathbf{x}_i \in \mathbb{R}^3$ and local radius $r_i \in \mathbb{R}^+$, and inlet flow rate q. The goal is to predict the pressure $\mathrm{P}(\mathbf{x})$ and the wall shear stress $\mathrm{WSS}(\mathbf{x})$ at arbitrary wall points while preserving morphology and flow features.

2.2 Architecture Overview

The model has two main components: (1) a transformer encoder that processes centerline geometry, and flow data, and (2) an anisotropic RBF decoder that reconstructs the continuous pressure field.

Transformer Encoder with 4D Fourier. Each point $\mathbf{c}_i = (x_i, y_i, z_i, r_i)$ is encoded with 4D Fourier positional embeddings:

$$\mathbf{e}(\mathbf{c}) = [\sin(2\pi\mathbf{B}\mathbf{c}); \cos(2\pi\mathbf{B}\mathbf{c})] \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{m\times 4}$ is the frequency basis matrix that maps 4D coordinates into Fourier space for positional encoding. The resulting embedding is concatenated with the raw coordinates and passed through a multilayer perceptron (MLP):

$$\mathbf{h}_i^{(0)} = \mathrm{MLP}\big(\big[\,\mathbf{e}(\mathbf{c}_i)\,;\,\mathbf{c}_i\,\big]\big)\,. \tag{2}$$

Learned positional embeddings are added before passing the sequence through transformer layers.

Flow Rate Conditioning. The inlet flow rate q modulates features using Feature-wise Linear Modulation (FiLM) [19]:

$$\gamma, \beta = \mathrm{MLP}(q) \tag{3}$$
$$\mathbf{h} \leftarrow \gamma \odot \mathbf{h} + \beta \tag{4}$$

This allows the network to adapt predictions to the boundary condition.

Table 1: Performance on synthetic single-vessel geometries (Test set). Results show ℓ_2 error for pressure, WSS, and their mean.

Method	Pressure	WSS	Mean
Low-Fidelity CFD	1.615	0.697	1.156
GNOT [21]	0.387	0.561	0.474
Transolver [22]	0.194	0.330	0.253
ONO [23]	0.205	0.290	0.248
Ours	**0.132**	**0.279**	**0.206**

Anisotropic RBF Decoder. The transformer outputs parameterize N anisotropic Gaussian kernels (RBF size) for each field $f \in \{\mathrm{P}, \mathrm{WSS}\}$, centered at $\mathbf{x}_{c,i}$:

$$\phi^{(f)}(\mathbf{x}, \mathbf{x}_{c,i}) = \exp\Big(-(\mathbf{x} - \mathbf{x}_{c,i})^T \mathbf{\Sigma}_i^{(f)-1} (\mathbf{x} - \mathbf{x}_{c,i})\Big). \tag{5}$$

For each field, covariance matrices are ensured positive semi-definite via Cholesky factorization $\mathbf{\Sigma}_i^{(f)\,-1} = \mathbf{L}_i^{(f)} \mathbf{L}_i^{(f)\top}$. With task-specific scalar weights $w_i^{(f)}$ predicted at each centerline point, the final reconstructions are

$$\mathrm{P}(\mathbf{x}) = \sum_{i=1}^{N} w_i^{(\mathrm{P})}\, \phi^{(\mathrm{P})}(\mathbf{x}, \mathbf{x}_{c,i}), \quad \mathrm{WSS}(\mathbf{x}) = \sum_{i=1}^{N} w_i^{(\mathrm{WSS})}\, \phi^{(\mathrm{WSS})}(\mathbf{x}, \mathbf{x}_{c,i}). \tag{6}$$

This lets the model adapt kernels for pressure and shear, using finer kernels in complex regions and broader ones in smoother segments.

3 Dataset

To construct the dataset, we generated synthetic coronary vessel geometries with controlled anatomical variations. The vessel length ranged from 40 mm to 70 mm, tapering ratios (defined as the ratio of outlet to inlet radius) varied between 0.6 and 0.8, and stenosis severity ranged from 30% to 70%. These parameters were used to generate the vessel centerline representation, consisting of coordinates and radius values, denoted as (x, y, z, r).

Using these centerlines, we constructed 3D surface and volume meshes and performed steady-state flow simulations in OpenFOAM [20]. Each case used physiologically relevant inlet pressures and flow rates. In total, 4000 single-vessel geometries were generated along with their respective flow fields. The dataset was then split into training and testing sets using a fixed 90%-10% partition, which was held constant across all experiments rather than being randomly resampled each time.

4 Experiments

4.1 Single Vessel Geometry Evaluation

We evaluate several variants of our model on the single-vessel synthetic dataset to assess the contribution of different components. In this experiment, we compare against neural operators, including GNOT [21], Transolver [22], and ONO [23], as well as a low-fidelity baseline. All models are trained to predict both pressure and wall shear stress fields. All models were trained for 20,000 epochs with 1.9M parameters, keeping other hyperparameters identical across experiments. See Appendix B.

Table 1 demonstrates that our model achieves substantially lower errors compared to baseline methods. Furthermore, Figure 2 shows that our predicted FFR aligns more closely with ground truth than competing approaches, highlighting both accuracy and robustness of our framework.

4.2 Computational Efficiency Analysis

To analyze the relationship between model configuration, performance, and computational cost, we performed two controlled experiments. In the first, we fixed the number of RBFs and varied the number of query sample points; in the second, we fixed the query sample size and varied the

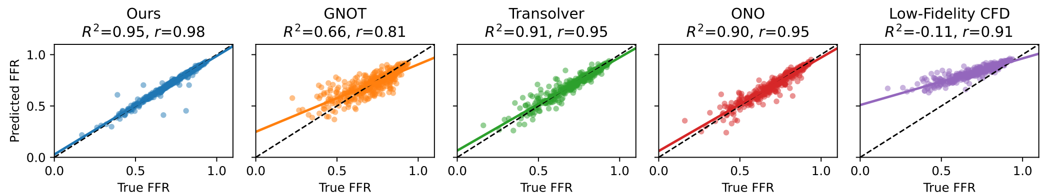

Figure 2: Predicted and true FFR across methods on testing set. Our model shows the strongest agreement ($R^2 = 0.95$, $r = 0.98$) compared to GNOT, Transolver, ONO, and low-fidelity CFD.

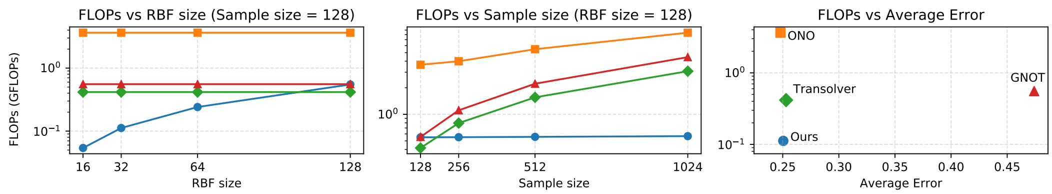

Figure 3: Comparison of computational cost and accuracy across models. (Left) Increasing the RBF size with fixed sample size (128) increases the FLOPs for our method, but the cost remains comparable to other methods at RBF size 128. (Middle) Increasing the number of query sample points does not increase the computational cost for our method, thanks to the shallow decoder design, whereas other models scale steeply. (Right) Our model is $\mathbf{1.72}\times$ faster than Transolver and $\mathbf{15.1}\times$ faster than ONO, while achieving lower or comparable errors.

number of RBFs. For each configuration, we measured the number of giga floating-point operations (GFLOPs) required for a single forward pass, which quantifies the computational cost, and compared these against baseline neural operator models. This setup allows us to isolate the impact of each factor, query sample size, and RBF size, on both computational complexity and predictive accuracy. As shown in Figure 3, our method achieves better accuracy per FLOP than neural-operator baselines. FLOPs increase with the number of RBFs, but at 128 bases the cost remains comparable across methods. Crucially, the shallow decoder makes inference cost (nearly) invariant to the number of query points, unlike baselines that scale steeply, introduce significant overhead in training time and inference cost as query counts increase. Overall, we retain competitive cost at practical RBF counts and achieve faster inference ($1.7\times$ vs. Transolver, $15\times$ vs. ONO, $2.3\times$ vs. GNOT).

4.3 The Effect of RBF Size

Figure 4 illustrates the relationship between the number of RBF bases and model accuracy. In general, increasing the RBF size provides the model with greater expressive capacity, which leads to reduced error up to a certain point. However, this improvement does not scale indefinitely: excessively large RBF counts (e.g., 512) cause performance to degrade, likely due to overparameterization and optimization difficulties under a fixed training budget. It shows a trade-off between accuracy and efficiency, suggesting that moderately sized bases achieve the best balance between representational power and trainability.

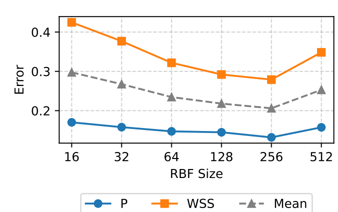

Figure 4: Test errors decrease with larger RBF counts, but rise at 512.

5 Conclusion

We present a Transformer-Anisotropic RBF framework for predicting wall pressure and shear stress in coronary arteries from centerline geometry, hemodynamic descriptors, and inlet flow rate. The model provides high-fidelity, continuous predictions at low cost, achieving lower ℓ_2 errors and up to $15\times$ faster inference than neural operator baselines. We also introduce a large-scale synthetic coronary hemodynamics dataset (4,000 single-vessel geometries with CFD-derived pressure and WSS), supporting robust training, reproducible benchmarking, and future work on multi-vessel anatomies. Currently, our model does not provide uncertainty estimates. In future work, both aleatoric uncertainty (from data variability) and epistemic uncertainty (from model limitations) could be incorporated using Bayesian approaches, ensembles, or probabilistic modeling. Together, the framework and dataset enable fast, anatomically aware hemodynamic assessment for both individual diagnosis and population-level screening.

References

[1] Salim S. Virani and et al. Heart disease and stroke statistics—2021 update: A report from the american heart association. *Circulation*, 143(8):e254–e743, 2021. doi: 10.1161/CIR.0000000000000950.

[2] Paul A. Tonino and et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *New England Journal of Medicine*, 360(3):213–224, 2009. doi: 10.1056/NEJMoa0807611.

[3] Nico H. J. Pijls, Bernard De Bruyne, K. Peels, P. H. Van Der Voort, H. J. Bonnier, J. Bartunek, and J. J. Koolen. Measurement of fractional flow reserve to assess the functional severity of coronary-artery stenoses. *New England Journal of Medicine*, 334(26):1703–1708, 1996. doi: 10.1056/NEJM199606273342604.

[4] Charles A. Taylor, Thomas A. Fonte, and John K. Min. Computed fractional flow reserve using coronary ct angiography: Physiological assessment of coronary stenoses at a fraction of the time and cost of invasive testing. *JACC: Cardiovascular Imaging*, 6(11):1209–1222, 2013. doi: 10.1016/j.jcmg.2013.08.002.

[5] Sashikumaar Sankaran, Hyun Jin Kim, Gi Won Choi, Irene E. Vignon-Clementel, and Charles A. Taylor. Patient-specific multiscale modeling of blood flow for coronary artery bypass graft surgery. *Annals of Biomedical Engineering*, 40(10):2228–2242, 2012. doi: 10.1007/s10439-012-0598-5.

[6] Georgios Kissas, Yibo Yang, Enrui Hwuang, Weiwen R. Witschey, John A. Detre, and Paris Perdikaris. Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 358:112623, 2020. doi: 10.1016/j.cma.2019.112623.

[7] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019. URL `https://api.semanticscholar.org/CorpusID:57379996`.

[8] Reza Akbarian Bafghi and Maziar Raissi. Pinns-tf2: Fast and user-friendly physics-informed neural networks in tensorflow v2. *ArXiv*, abs/2311.03626, 2023. URL `https://api.semanticscholar.org/CorpusID:265043331`.

[9] Nikola B. Kovachki, Zong-Yi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *J. Mach. Learn. Res.*, 24:89:1–89:97, 2023. URL `https://api.semanticscholar.org/CorpusID:259149906`.

[10] Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *J. Comput. Phys.*, 449:110768, 2020. URL `https://api.semanticscholar.org/CorpusID:220845655`.

[11] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne J. Bergen, Carla P. Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun K. Manrai, Debora S. Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Velickovic, Max Welling, Linfeng Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial intelligence. *Nature*, 620:47–60, 2023. URL `https://api.semanticscholar.org/CorpusID:260384616`.

[12] Zong-Yi Li, Nikola B. Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *ArXiv*, abs/2010.08895, 2020. URL `https://api.semanticscholar.org/CorpusID:224705257`.

[13] Zong-Yi Li, Nikola B. Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *ArXiv*, abs/2003.03485, 2020. URL `https://api.semanticscholar.org/CorpusID:211838009`.

[14] Zong-Yi Li, Nikola B. Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3d pdes. *ArXiv*, abs/2309.00583, 2023. URL `https://api.semanticscholar.org/CorpusID:261494027`.

[15] Georgios Kissas, Jacob H. Seidman, Leonardo Ferreira Guilhoto, Victor M. Preciado, George Pappas, and Paris Perdikaris. Learning operators with coupled attention. *ArXiv*, abs/2201.01032, 2022. URL `https://api.semanticscholar.org/CorpusID:245668802`.

[16] Xinliang Liu, Bo Xu, Shuhao Cao, and Lei Zhang. Mitigating spectral bias for the multiscale operator learning. *J. Comput. Phys.*, 506:112944, 2022. URL https://api.semanticscholar.org/CorpusID:260316030.

[17] Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. Dpot: Auto-regressive denoising operator transformer for large-scale pde pre-training. *ArXiv*, abs/2403.03542, 2024. URL https://api.semanticscholar.org/CorpusID:268253199.

[18] Benjamin Shih, Ahmad Peyvan, Zhongqiang Zhang, and George Em Karniadakis. Transformers as neural operators for solutions of differential equations with finite regularity. *ArXiv*, abs/2405.19166, 2024. URL https://api.semanticscholar.org/CorpusID:270095024.

[19] Ethan Pérez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[20] H. Jasak. Openfoam: Open source cfd in research and industry. *International journal of naval architecture and ocean engineering*, 1(2):89–94, 2009. doi: https://doi.org/10.2478/IJNAOE-2013-0011.

[21] Zhongkai Hao, Chengyang Ying, Zhengyi Wang, Hang Su, Yinpeng Dong, Songming Liu, Ze Cheng, Jun Zhu, and Jian Song. Gnot: A general neural operator transformer for operator learning. *ArXiv*, abs/2302.14376, 2023. URL https://api.semanticscholar.org/CorpusID:257232579.

[22] Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *ArXiv*, abs/2402.02366, 2024. URL https://api.semanticscholar.org/CorpusID:267411758.

[23] Zipeng Xiao, Zhongkai Hao, Bokai Lin, Zhijie Deng, and Hang Su. Improved operator learning by orthogonal attention. *ArXiv*, abs/2310.12487, 2023. URL https://api.semanticscholar.org/CorpusID:264306159.

A Overview

This appendix provides additional details supporting Section 4. We describe training hyperparameters and model configuration, outline the low-fidelity baseline, define key evaluation metrics, and present supplementary quantitative results, including extended training-set performance.

B Experiments

Hyperparameters. We optimize with Adam (learning rate 5×10^{-4}, weight decay 5×10^{-4}). The schedule uses a $1{,}000$-epoch linear warmup from $10^{-8}\times$LR to full LR, followed by cosine annealing to 2.5×10^{-6} for the remaining epochs, for a total of $20{,}000$ epochs. We use batch size 512, train on $3{,}600$ cases and validate on 400, and fix the random seed to 1234 for reproducibility. For evaluation, we report results from the checkpoint that achieves the best mean accuracy across validation epochs.

Transformer Encoder Details. We use a transformer encoder with 4 layers, 1 attention head, model dimension 256, and feedforward dimension 256, with dropout 0.1. Tokens are built from 4D inputs (x, y, z, r) using Fourier embeddings with a hidden dimension of 48 that are projected to the model dimension via a linear layer. We employ learned positional encodings over the token sequence.

Fractional Flow Reserve (FFR). FFR is a clinically important hemodynamic index used to assess the functional significance of coronary artery stenoses. It is defined as the ratio between the mean distal coronary pressure (P_d) and the mean aortic pressure (P_a) under conditions of maximal hyperemia:

$$\mathrm{FFR} = \frac{P_d}{P_a}. \tag{7}$$

An FFR value below 0.80 is typically considered indicative of flow-limiting stenosis and is widely used to guide revascularization decisions.

Low Fidality Baseline. We compute 1D pressure along the fitted centerline using Poiseuille's law. The axial pressure gradient is obtained from the volumetric flow rate and vessel radius, integrated along the centerline to yield pressure.

Training Performance. As shown in Table 2, our method consistently outperforms all baselines on the training set, achieving substantially lower ℓ_2 errors in both pressure and WSS. These improvements mirror the validation results in Table 1, where our model also achieves the lowest errors across metrics. Together, these findings highlight not only the strong representational capacity of our approach during training but also its ability to generalize robustly to unseen cases.

Table 2: Performance on synthetic single-vessel geometries (Training set). Results show ℓ_2 error for pressure, WSS, and their mean.

Method	Pressure	WSS	Mean
GNOT [21]	0.253	0.426	0.340
Transolver [22]	0.176	0.314	0.245
ONO [23]	0.173	0.255	0.214
Ours	**0.050**	**0.200**	**0.125**