PoseBridgeNet: Learning a Pose-Evolving Schrödinger Bridge for Multi-Organs Segmentation in Prenatal Volumetric Ultrasound

Shushen Gan[1], Yongqian Wu[2], JinQiao Duan[3,4], Jiahong Liang[1],Chengdong Deng[2],Wanqin Zhang[2],Xiaojing Zhou[2], Yixin Zhu[2], Yu Shi[2*], Yi Gao[1,3,5*]

1 School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen 518060, China
gaoyi@szu.edu.cn
2 Department of Ultrasonography, Peking University Shenzhen Hospital, China
shiyu@pkuszh.com
3 Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems, Dongguan 523000, China
4 Department of Mathematics and Department of Physics, Great Bay University, Dongguan, China
5 Marshall Laboratory of Biomedical Engineering, Shenzhen 518060, China

Abstract. Accurate thoracoabdominal multi-organ segmentation of the fetal from mid-trimester volumetric ultrasound is crucial for prenatal screening and assessment of fetal development. Over half of severe structural anomalies can be detected through such examinations. However, this task is particularly challenging due to the inherent low contrast and ill-defined boundaries in volumetric ultrasound, which are compounded by significant feature distribution shifts caused by variations in fetal pose. To address this, we propose PoseBridgeNet, a novel framework for segmenting six key organs in this challenging setting. Our core innovation is the integration of a Pose-Evolving Schrödinger Bridge (PSB) within the bottleneck of the segmentation model. This module formulates the alignment of feature distributions across different poses as an entropy-regularized optimal transport problem. By explicitly simulating the dynamic evolution of features across pose variations via a learnable stochastic differential equation, the model gains robustness against pose-induced feature shifts. Furthermore, we design a Mixture-of-Experts (MoE) decoder that integrates the stochastic features from the Pose-Evolving Schrödinger Bridge: it consists of one shared expert capturing cross-pose anatomical commonalities and multiple specialized experts modeling organ- and pose-specific variations, enabling joint optimization of pose-aware feature alignment and segmentation. Evaluated on an in-house dataset of fetal volumetric ultrasound, PoseBridgeNet outperforms existing methods, achieving an average DSC of 85.52% , an IoU of 75.44% and a HD of 3.10 mm. Ablation studies further validate the effectiveness of each proposed component. The code and data are available at https://github.com/GanSusan/PoseBridgeNet.git.

Keywords: Medical image segmentation · Fetal ultrasound · Schrödinger bridge.

1 Introduction

Mid-trimester fetal thoracoabdominal ultrasound is essential for prenatal screening, detecting over half of severe structural anomalies such as bilateral renal agenesis and single ventricle defect. Volumetric probes now acquire three-dimensional anatomical data covering all key thoracoabdominal organs in a single scan, enabling automated analysis via deep learning-based multi-organ segmentation to assist diagnosis and fetal growth assessment.

However, automatic segmentation faces significant technical challenges. As illustrated in Fig.1(a)-(b) and (e), inherent imaging limitations include low inter-organ contrast, along with low resolution and ill-defined boundaries. More critically, fetal pose variation exacerbates these issues. As shown in Fig.1(c)-(d), postural changes within the uterus substantially alter organs' spatial arrangement, projected morphologies, and mutual occlusion relationships, inducing systematic feature distribution shifts. Fig.1(f) demonstrates this with the liver: its grayscale disparity across different poses approaches that between distinct organs. This pose-induced feature distribution shift undermines performance consistency, limiting the generalization of existing methods.

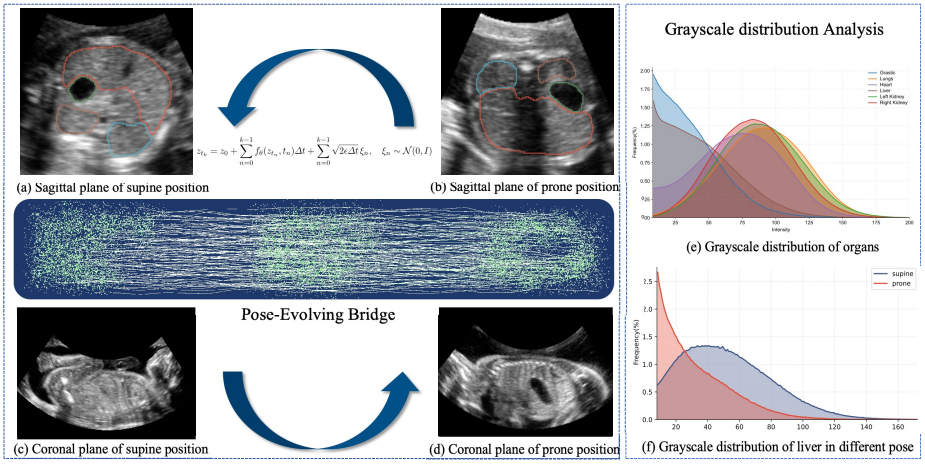

Fig. 1. Overview of fetal thoracoabdominal volumetric ultrasound images and Pose-Evolving Bridge. (a,b) Abdominal organs in supine and prone positions; (c,d) Coronal views in supine and prone positions.(e) Grayscale distributions of all organs; (f) Liver grayscale distribution in supine vs. prone positions.

Existing fetal volumetric ultrasound studies have explored images of the first-trimester [1], head [2], and femur [3], but research on mid-trimester thoraco-

abdominal images remains scarce. Although prior work have recognized the importance of pose modeling in prenatal ultrasound, proposing spatial normalization preprocessing frameworks[4] or deep learning-based pose estimation algorithms [5], these methods have not integrated pose modeling with the segmentation model to address the impact of pose-induced feature distribution shifts on segmentation performance.

Recently, the Schrödinger bridge framework has gained traction in computer vision. Rooted in entropy-regularized optimal transport, it constructs learnable stochastic interpolations between distributions, demonstrating strong performance in image generation [6–8], score-based dynamics [9, 10], and cross-modal translation [11–13]. This provides a principled mathematical framework for modeling feature shifts induced by fetal pose variations. Meanwhile, Mixture-of-Experts (MoE) architectures enhance representational capacity for complex data via dynamic gating [14], with recent medical segmentation studies integrating MoE into U-Net [15] and SAM [16] for improved feature fusion [17].

We propose PoseBridgeNet, a multi-organ segmentation network for mid-trimester fetal thoracoabdominal volumetric ultrasound. Built upon U-Net [18], our model introduces two core innovations: a pose-evolving Schrödinger bridge in the bottleneck layer that explicitly models pose-induced feature shifts, and a MoE decoder that leverages the bridge's stochastic features for joint optimization of the Schrödinger bridge and segmentation tasks. Our main contributions are:

1. The first multi-organ segmentation model for mid-trimester fetal thoracoabdominal volumetric ultrasound, addressing an underexplored yet clinically critical application.
2. A pose-evolving Schrödinger bridge in the bottleneck layer that explicitly models feature distribution shifts caused by fetal pose variations via a learnable stochastic process, conferring inherent robustness to pose changes.
3. A MoE decoder that effectively integrates stochastic features from the pose-evolving bridge and enables joint optimization of the Schrödinger bridge formulation and segmentation, achieving appreciable performance gains on challenging ultrasound data.

2 Method

2.1 The Pose-Evolving Schrödinger Bridge

Problem Formulation We consider the volumetric ultrasound images of a fetus under different poses as a continuous family of distributions, where variations in the pose parameter θ induce a continuous shift in the feature distribution F_θ. For any two poses θ_s and θ_t, we aim to construct a stochastic process that transports features from the distribution F_{θ_s} to F_{θ_t} such that the path of this process follows the continuous evolution of anatomical structures. This objective aligns with the Schrödinger Bridge framework—an entropy-regularized optimal transport model that seeks the most likely stochastic evolution path given prescribed boundary distributions.

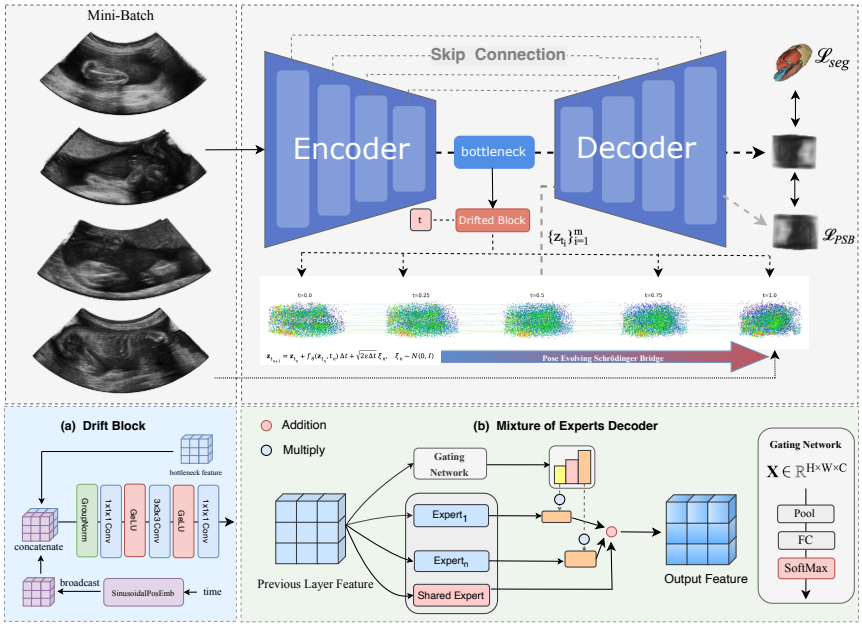

Fig. 2. Overall framework of our proposed PoseBridgeNet

To construct the boundary distributions for the evolution path during training, for each input $\mathbf{X}_i$ (corresponding to pose θ_i), we denote its bottleneck feature as $z_0 \in R^{C\times H\times W}$. From the current batch, we randomly select another sample $\mathbf{X}_j$ (corresponding to pose θ_j) whose bottleneck feature is denoted as z_1. We treat z_0 as an instance drawn from the source distribution P_S and z_1 as an instance drawn from the target distribution P_T. By guiding the model to learn the Schrödinger bridge from P_S to P_T, the model acquires the ability to robustly model features across different poses.

Stochastic Evolution Process and Its Parameterization Taking z_0 as the initial state, we define a stochastic process $\{z_t\}_{t\in[0,1]}$ that satisfies the Itô stochastic differential equation (SDE):

$$\mathrm{d}z_t = f(z_t, t)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t \tag{1}$$

where $f : R^C \times [0,1] \to R^C$ denotes the learnable drift function, ε is a fixed diffusion coefficient, and W_t represents standard Brownian motion. The terminal distribution z_1 of this SDE is required to approximate the target distribution P_T, which is the objective of the Schrödinger bridge problem. The drift function, which takes the time t and the current state z_t as inputs, is implemented by a drift block consisting of two $1\times1\times1$ convolutional layers and one $3\times3\times3$ convolutional layer (Fig. 2(a)).

$$f(z_t, t) = DriftBlock\left(GroupNorm\left([z_t; \gamma(t)]\right)\right) \tag{2}$$

Here, $\gamma(t)$ is the sinusoidal positional embedding of time t. This design enables the drift function to dynamically adjust features according to the evolution time step, guiding a smooth transition of features while preserving anatomical consistency.

Numerical Simulation and Consistency Learning We simulate the stochastic differential equation using the Euler–Maruyama method. Let the time interval $[0, 1]$ be discretized into N steps with step size $\Delta t = 1/N$, and denote the discrete time points as $t_n = n\Delta t$ for $n = 0, 1, \ldots, N$. The discretized form of the differential equation can be expressed as follows: for any time t_k, the solution is given by:

$$z_{t_k} = z_0 + \sum_{n=0}^{k-1} f_\theta(z_{t_n}, t_n)\Delta t + \sum_{n=0}^{k-1} \sqrt{2\epsilon\Delta t}\,\xi_n, \quad \xi_n \sim \mathcal{N}(0, I) \tag{3}$$

Eq. (3) yields the entire discretized trajectory $\{z_{t_0}, z_{t_1}, \ldots, z_{t_N}\}$ with $z_{t_0} = z_0$ and $z_{t_N} = \hat{z}_1$ denoting the terminal feature.

To obtain diverse intermediate features under different poses, during training we independently sample m time points $\{t_i\}_{i=1}^m \sim \mathcal{U}(0, 1)$ along the evolution path and compute the corresponding features $\{z_{t_i}\}_{i=1}^m$ using Eq. (3). These features, together with the original bottleneck feature z_0, are fed into the decoder to produce the corresponding segmentation probability maps P_0 and $\{P_{t_i}\}_{i=1}^m$, respectively.

$$\mathcal{L}_{PSB} = \lambda \cdot \frac{1}{m} \sum_{i=1}^{m} D_{\mathrm{KL}}(P_0 \| P_{t_i}) + \gamma \|\hat{z}_1 - z_1\|_2^2, \tag{4}$$

where $\lambda > 0$, $\gamma > 0$ are hyperparameters. The first term implicitly guides the Schrödinger bridge evolution path to maintain segmentation consistency, while the second term steers the terminal state toward the target feature associated with the Schrödinger bridge's boundary condition.

2.2 Mixture-of-Experts Decoder

To effectively leverage the stochastic features from the Pose-Evolving Schrödinger Bridge while jointly optimizing the Schrödinger bridge and segmentation, we equip the decoder with a MoE module (Fig. 2(b)). It comprises one shared expert capturing cross-pose anatomical commonalities and N specialized experts modeling organ- and pose-specific variations. Given an input $\mathbf{X} \in R^{H \times W \times C}$, the gating network applies global average pooling to obtain a global context vector $\mathbf{X}_{global} \in R^C$, then computes expert weights $\mathbf{g} = [g_1, g_2, \ldots, g_N]^\top$ via linear mapping and softmax:

$$g_i = \frac{\exp\left(\mathbf{w}_i^\top \mathbf{X}_{global} + b_i\right)}{\sum_{j=1}^{N} \exp\left(\mathbf{w}_j^\top \mathbf{X}_{global} + b_j\right)} \tag{5}$$

where $\mathbf{w}_i \in R^C$ and $b_i \in R$ are learnable parameters of the gating network. The shared and N specialized experts are implemented as $3 \times 3 \times 3$ convolutional layers, producing $F_{shared} \in R^{H \times W \times C}$ and $\{F_i \in R^{H \times W \times C}\}_{i=1}^{N}$, respectively. The final output Y combines shared features and the weighted sum of specialized features:

$$Y = F_{shared} + \sum_{i=1}^{N} g_i F_i \tag{6}$$

2.3 Optimization Objective

We formulate an end-to-end trainable objective integrating multi-organ segmentation supervision with pose evolution. The final loss is a sum of segmentation loss $\mathcal{L}_{Seg}$, pose-evolving bridge loss $\mathcal{L}_{PSB}$ and L2 regularization:

$$\mathcal{L} = \mathcal{L}_{Seg} + \mathcal{L}_{PSB} + \alpha \|\mathbf{W}\|_2^2 \tag{7}$$

where $\alpha > 0$ is a hyperparameter and $\|\mathbf{W}\|_2^2$ is L2 weight decay.

3 Experiments and Results

3.1 Implementation Details and Datasets

Implementation Details We implemented our model using PyTorch 2.6.0 and MONAI 1.4.0 on two NVIDIA A6000 GPUs. All input volumes were resampled to $128 \times 128 \times 128$ voxels with isotropic spacing. The network was optimized using AdamW with an initial learning rate of 1×10^{-3}, and a weight decay of 1×10^{-5}. Early stopping was applied if the validation metric did not improve for 10 consecutive epochs. Five-fold cross-validation was performed to ensure reliable evaluation.

Datasets and Evaluation Metrics We collected fetal volumetric ultrasound scans (21–24 weeks) from 80 subjects using a GE Voluson scanner with a RAB6-D probe. All volumes were manually annotated by two experts with over 10 years of experience for six organs: heart, lungs, liver, stomach, left kidney, and right kidney. Segmentation performance was evaluated using Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and 95th-percentile Hausdorff Distance (HD).

3.2 Comparison with Other Methods

Quantitative Comparison We compared our method with representative 3D segmentation networks [18–22]. As shown in Table 1, our approach achieves state-of-the-art performance, with mean DSC of 85.52%, mean IoU of 75.44%, and mean HD of 3.10 mm. Compared to the second-best method (VNet [21]), ours improves DSC by 0.6% (from 84.92% to 85.52%) and IoU by 0.8% (from

Table 1. Comparison of different methods on thoracic organs and overall performance. Overall denotes the average metric over six organs. All metrics are reported as mean values across folds (DSC and IoU in %, HD in mm). Standard deviations are omitted for brevity. The best results are highlighted in bold. * indicates statistical significance between the best and second-best results (Wilcoxon signed-rank test, $p < 0.05$)

Method	Lungs			Heart			Overall (6 organs)		
	DSC	IoU	HD	DSC	IoU	HD	DSC	IoU	HD
3D UNet[18]	84.92	74.47	4.93	84.60	74.05	5.72	83.15	72.16	5.76
VNet[21]	86.28	76.46	3.60	86.26	76.43	4.12	84.92	74.64	4.01
Att-UNet[19]	86.09	76.53	6.91	85.92	76.14	6.12	84.20	73.83	7.76
SegResNet[22]	86.11	76.19	4.15	86.33	76.54	4.54	84.65	74.30	4.86
SwinUNETR[20]	83.70	72.98	6.87	84.57	73.85	7.14	82.59	71.48	6.86
Ours	**87.25***	**77.87***	**2.78***	**86.49***	**76.79***	**2.91***	**85.52***	**75.44***	**3.10***

Table 2. Comparison of Different Methods on Abdominal Organs. All metrics are reported as mean values across folds (DSC and IoU in %, HD in mm). Standard deviations are omitted for brevity. The best results are highlighted in bold.* indicates statistical significance between the best and second-best results (Wilcoxon signed-rank test, $p < 0.05$)

	Stomach			Liver			Left Kidney			Right Kidney		
Method	DSC	IoU	HD	DSC	IoU	HD	DSC	IoU	HD	DSC	IoU	HD
3D UNet	87.60	78.57	4.73	84.86	74.41	5.71	79.91	67.62	5.77	77.01	63.86	6.99
VNet	**89.53**	81.47	1.57	85.61	75.46	4.12	82.16	70.64	4.52	79.70	67.43	5.86
Att-UNet	88.24	79.78	3.80	85.13	75.09	6.12	81.24	69.57	10.27	78.56	65.90	12.18
SegResNet	88.98	80.70	2.36	85.54	75.43	4.53	81.29	69.72	6.08	79.67	67.22	6.03
SwinUNETR	88.29	79.68	4.33	83.55	72.58	7.15	78.60	66.03	9.01	76.84	63.79	7.79
Ours	87.87	**81.54***	**1.50**	**87.56***	**78.36***	**3.97***	**82.80***	**71.56***	**3.68**	**81.13***	**69.10***	**4.00***

74.64% to 75.44%), while reducing HD by 22.7% (from 4.01 mm to 3.10 mm), demonstrating superior region overlap and boundary accuracy. Notably, as shown in Table 2, the liver, which is the organ most susceptible to pose variations due to acoustic shadowing from the spine (Fig. 1), achieves the largest improvement among abdominal organs, with DSC increasing by 1.95% (from 85.61% to 87.56%), IoU by 2.9% (from 75.46% to 78.36%), and HD decreasing by 3% (from 4.12 mm to 3.97 mm), validating the effectiveness of our method in modeling pose-induced feature shifts. We further performed the Wilcoxon signed-rank test between the best and second-best methods, and found that all improvements were statistically significant ($p < 0.05$), except for the HD95 on the stomach and left kidney.

Qualitative Comparison Fig. 3 presents a qualitative comparison of segmentation results produced by different methods. Our model exhibits superior robustness and modeling capacity. As shown in the first and second rows of Fig. 3, competing methods struggle to distinguish between the left and right kidney regions. Additionally, their segmentation of other organs exhibits suboptimal smoothness and false positives. For particularly challenging cases, exempli-

fied in the third row of Fig. 3, some methods generate substantial false positives, severely degraded in segmentation. Our model, however, continues to produce reliable segmentations under such difficult conditions.

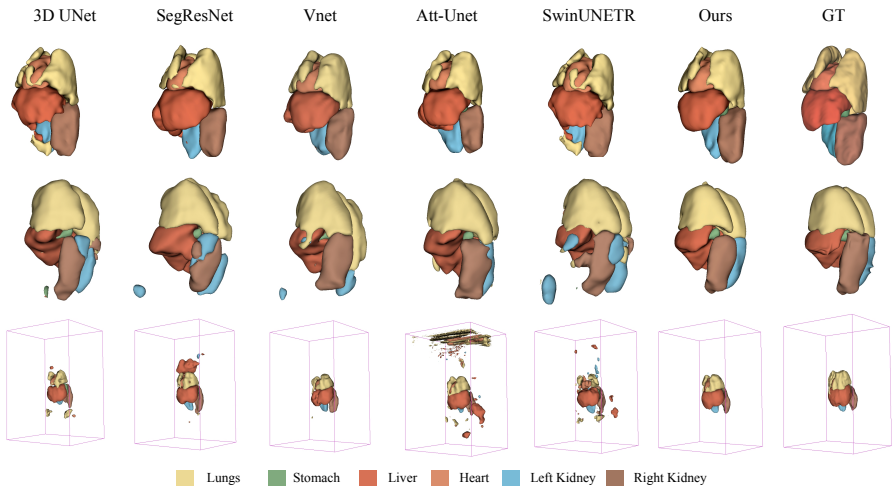

Fig. 3. 3D visualization of segmentation results from different models.

3.3 Ablation Study

We performed ablation studies on PoseBridgeNet, with results summarized in Table 3. All metrics are computed by first averaging over five folds for each organ and then across all six organs. Using a 3D U-Net as the base model (DSC 83.15%, IoU 72.16%, HD 5.76 mm), adding the Pose-Evolving Schrödinger Bridge (PSB) improved DSC by 0.69%, IoU by 1.01%, and reduced HD by 1.27 mm by explicitly modeling feature dynamics across poses to mitigate feature shifts. Adding only the MoE decoder achieved increases of 2.27% in DSC, 3.18% in IoU, and a reduction of 1.46 mm in HD, validating its ability to handle diverse multi-organ morphologies under varying poses. Integrating both modules yielded the best performance: DSC of 85.52%, IoU of 76.44%, and HD of 3.10 mm. This synergy demonstrates that PSB generates rich pose-evolved feature representations while MoE effectively leverages them for multi-organ segmentation, jointly enhancing robustness on challenging fetal thoracoabdominal ultrasound data.

4 Conclusion

We propose PoseBridgeNet for mid-trimester fetal thoracoabdominal multi-organ segmentation. To address feature shift caused by fetal pose variations, we in-

Table 3. Ablation study on PoseBridgeNet. Results are reported as mean ± standard deviation. The best results are highlighted in bold.

Method	DSC (%)	IoU (%)	HD (mm)
Base model	83.15 ± 9.33	72.16 ± 12.66	5.76 ± 9.95
Base model + PSB	83.84 ± 9.29	73.17 ± 12.61	4.49 ± 6.71
Base model + MoE	85.42 ± 7.96	75.34 ± 11.38	4.30 ± 7.39
Base model + PSB + MoE	$\mathbf{85.52 \pm 4.59}$	$\mathbf{75.44 \pm 7.74}$	$\mathbf{3.10 \pm 3.68}$

troduce a Schrödinger bridge-based stochastic process in the bottleneck layer that explicitly models feature evolution via a learnable SDE. A MoE decoder integrates shared and specialized experts to fuse stochastic features for segmentation. Experiments show our method outperforms existing approaches on the DSC, IoU, and HD, demonstrating robustness to pose-induced shifts. Ablation studies confirm the complementary effectiveness of both modules.

Acknowledgments. This work was partially supported by the Shenzhen Science and Technology Program (Grants JCYJ20250604182556073 and JCYJ2024 - 0806134133002), the Shenzhen Medical Research Fund (C2501029), and the Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems (2024B1212010004).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Yang, X., Yu, L., Li, S., Wen, H., Luo, D., Bian, C., Qin, J., Ni, D., Heng, P.-A.: Towards Automated Semantic Segmentation in Prenatal Volumetric Ultrasound. IEEE Trans. Med. Imaging. **38**, 180–193 (2019)
2. Yang, X., Wang, X., Wang, Y., Dou, H., Li, S., Wen, H., Lin, Y., Heng, P.-A., Ni, D.: Hybrid attention for automatic segmentation of whole fetal head in prenatal ultrasound volumes. Computer Methods and Programs in Biomedicine. **194**, 105519 (2020)
3. Wang, X., Yang, X., Dou, H., Li, S., Heng, P.-A., Ni, D.: Joint Segmentation and Landmark Localization of Fetal Femur in Ultrasound Volumes. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 1–5. IEEE, Chicago, IL, USA (2019)
4. Raynaud, C. et al.: Multi-organ Detection in 3D Fetal Ultrasound with Machine Learning. In: Cardoso, M. et al. (eds.) Fetal, Infant and Ophthalmic Medical Image Analysis. OMIA FIFI 2017, LNCS, vol. 10554. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67561-9_7
5. Yang, X., Shi, W., Dou, H., Qian, J., Wang, Y., Xue, W., Li, S., Ni, D., Heng, P.-A.: FetusMap: Fetal Pose Estimation in 3D Ultrasound. In: Shen, D. et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019, LNCS, vol. 11768. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_32
6. Peluchetti, S.: Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. Journal of Machine Learning Research **24**(374), 1–51 (2023)

7. Bortoli, V.D., Thornton, J., Heng, J., Doucet, A.: Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In: Ranzato, M., Beygelzimer, A. (eds.) Advances in Neural Information Processing Systems, vol. 34, pp. 17695–17709. Curran Associates, Inc. (2021).
8. Wang, G., Jiao, Y., Xu, Q., Wang, Y., Yang, C.: Deep Generative Learning via Schrödinger Bridge. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning, PMLR, vol. 139, pp. 10794–10804 (2021).
9. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion Schrödinger bridge with applications to score-based generative modeling. In: Advances in Neural Information Processing Systems 34 (NeurIPS 2021), pp. 17695–17709. Curran Associates, Inc., Red Hook, NY, USA (2021)
10. Somnath, V.R., Pariset, M., Hsieh, Y.-P., Martinez, M.R., Krause, A., Bunne, C.: Aligned Diffusion Schrödinger Bridges. In: Evans, R.J., Shpitser, I. (eds.) Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence, PMLR, vol. 216, pp. 1985–1995 (2023).
11. Teimouri, R., Kersten-Oertel, M., Xiao, Y.: CT-Based Brain Ventricle Segmentation via Diffusion Schrödinger Bridge without target domain ground truths. In: Linguraru, M.G. et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. MICCAI 2024, LNCS, vol. 15008. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72111-3_13
12. Bakr, E.M., Zhao, L., Hu, V.T., Cord, M., Perez, P., Elhoseiny, M.: ToddlerDiffusion: Interactive Structured Image Generation with Cascaded Schrödinger Bridge. In: The Thirteenth International Conference on Learning Representations (2025).
13. Jeong, Y., Cho, H., Yoon, K.-J.: Towards Robust Event-Based Networks for Nighttime via Unpaired Day-to-Night Event Translation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., and Varol, G. (eds.) Computer Vision – ECCV 2024, pp. 286–306. Springer Nature Switzerland, Cham (2025)
14. Wang, J., Yoshie, O., Ieiri, Y.: Conducting patch contrastive learning with mixture of experts on mixed datasets for medical image segmentation. Neural Comput & Applic. **37**, 14189–14216 (2025)
15. Luo, Z., Zhu, X., Zhang, L., Sun, B.: Rethinking U-Net: task-adaptive mixture of skip connections for enhanced medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, No. 6, pp. 5874–5882. AAAI Press (2025)
16. Li, R., Wu, L., Gu, J., Xu, Q., Chen, W., Cai, X., Bu, J.: MoE-SAM: Enhancing SAM for Medical Image Segmentation with Mixture-of-Experts. In: Gee, J.C. et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. MICCAI 2025, LNCS, vol. 15961. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-04937-7_35
17. Ou, Y., Yuan, Y., Huang, X., Wong, S.T.C., Volpi, J., Wang, J.Z., Wong, K.: Patcher: Patch Transformers with Mixture of Experts for Precise Medical Image Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022, LNCS, vol. 13435. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_46
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015, LNCS, vol. 9351. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

19. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.C.H., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S.G., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999 (2018).
20. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021, LNCS, pp. 272–284. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08999-2_22
21. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016). https://doi.org/10.1109/3DV.2016.79
22. Myronenko, A.: 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018, LNCS, vol. 11384. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_28