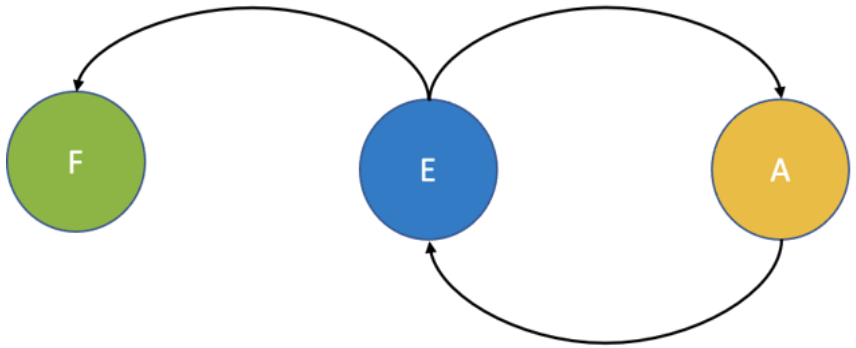
Outline

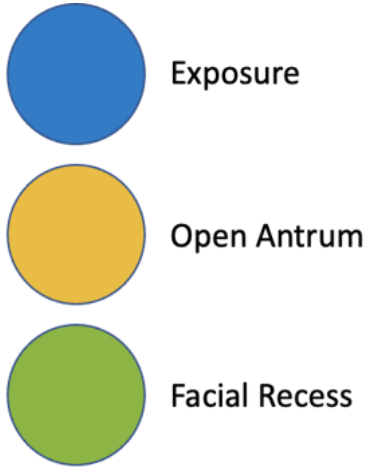
- Project Overview (importance)
- Deliverables Recap
- Dependencies Check
- Current progress
  - Preprocessed Mastoid dataset
  - Existing methods benchmarking
  - DL framework for Mastoid dataset (documentations, code management)
- Challenges & Problems
- Updated plans (Gannt chart, Milestones)
- Future works

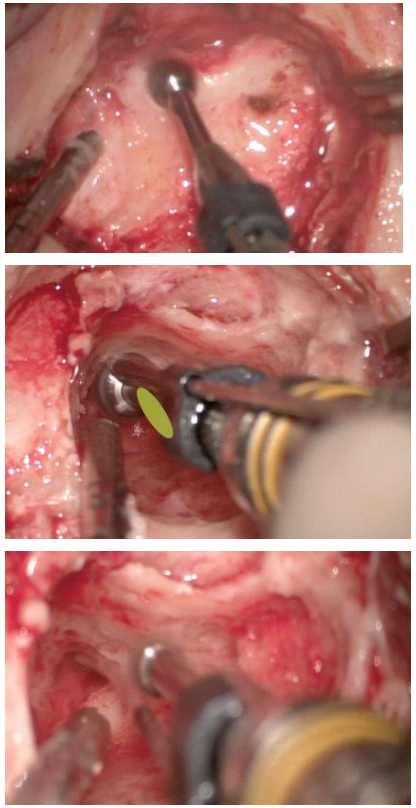
Project Overview:

- Our task:

  Mastoidectomy Surgical Video Segmentation using DL method.

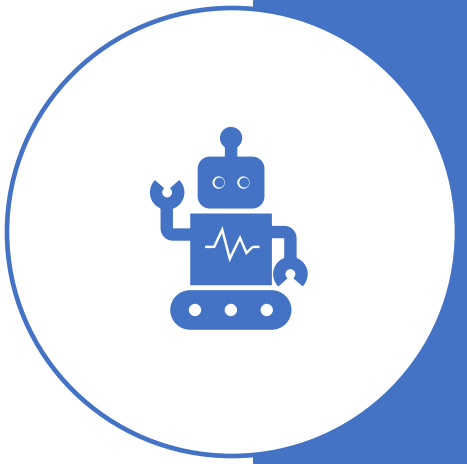






Importance:

- Importance of an automatic real-time Surgical Phase Detection:
  - Operating room monitoring and scheduling.
  - Surgical Video database indexing and analysis.
  - Foundation of a context-aware system and other downstream applications.



Updated Deliverables:

Deliverables:	Old	Updated	Status
Minimum	1. Dataset generation. 2. Benchmark three existing methods that are designed for a similar task (Cholecystectomy).	1. Same 2. Same **3. A well-documented benchmarking code base that can be easily reused to test out any model for our dataset.**	Completed
Expected	1. Experiment results analysis of the weakness of benchmarked models. 2. Redesign a new model that achieved 90% accuracy in the test set.	1. Same 2. **Redesign a new model that achieved 85% accuracy, 80% recall, and precision.**	In progress
Maximum	Conference paper	Same	During the summer

Dependencies:

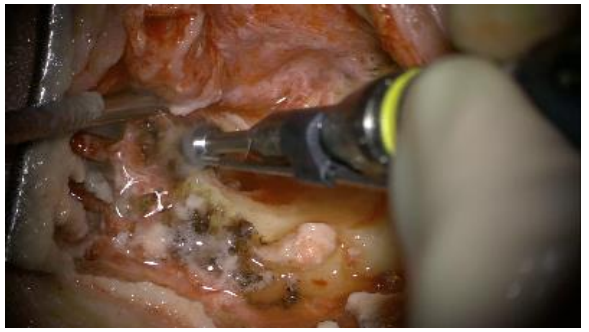
Main Dependencies	Sub Dependencies	Contact	Expected Date	Status	Alternative solution
Dataset	Data Generation	Dr. Danielle Trakimas	04/01	Complete	N/A
	Annotation Protocol	Dr. Danielle Trakimas	02/18	Complete	N/A
	Data Annotation	Dr. Danielle Trakimas	03/17	Ongoing	N/A
	IRB Training	Dr. Danielle Trakimas	02/11	Complete	N/A
	IRB Amendment	Dr. Danielle Trakimas	02/25	Complete	Use the safe desktop to do the preprocessing of the video, and onedrive streaming will be the alternative solution to address the failure of the IRB amendment
Computational Resources	GPU	Max Li	02/18	Complete	Use the online GPU resource such as Amazon cloud or Colab(Need to get the budget from mentors)
	Server Remote Access	Anton Deguet	02/18	Complete	Set up the computer in a physically available environment, and we need to use that computer to finish the project
Existing Framework & Public Dataset	Framework	Max Li	02/11	Complete	Implement and reproduce the frameworks based on the paper by ourselves using PyTorch
	Laparoscopic Public Dataset (Cholec80)	Max Li	02/11	Complete	Find Another available public dataset
Clinical Advice	Clinical Advice	Dr. Danielle Trakimas	/	Ongoing	Need to find another expert to provide clinical advice

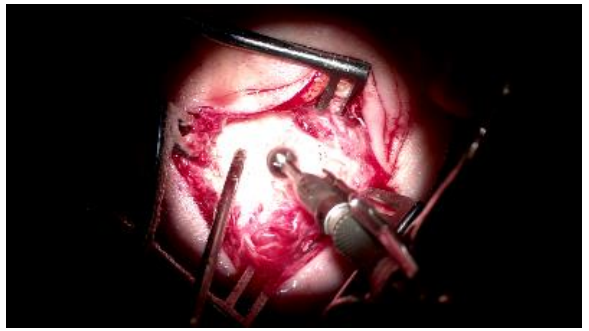
Annotated videos:

- Current: 12
- Expected: 15

Current Progress: Preprocessed Mastoid Dataset

- **Overview**:
  - 12 Mastoidectomy Surgical Vidoes with Surgical Phase Annotations
  - ~25-60min per video with 30 fps
  - ~1,000,000 frames in total
  - 3 labels: Expose, Open Antrum, Facial Recess
- **Preprocessing**:
  - Removed corrupted video frames
  - Downsampled to 225 X 400





Current Progress: Existing Methods Benchmarking -- Overall Results

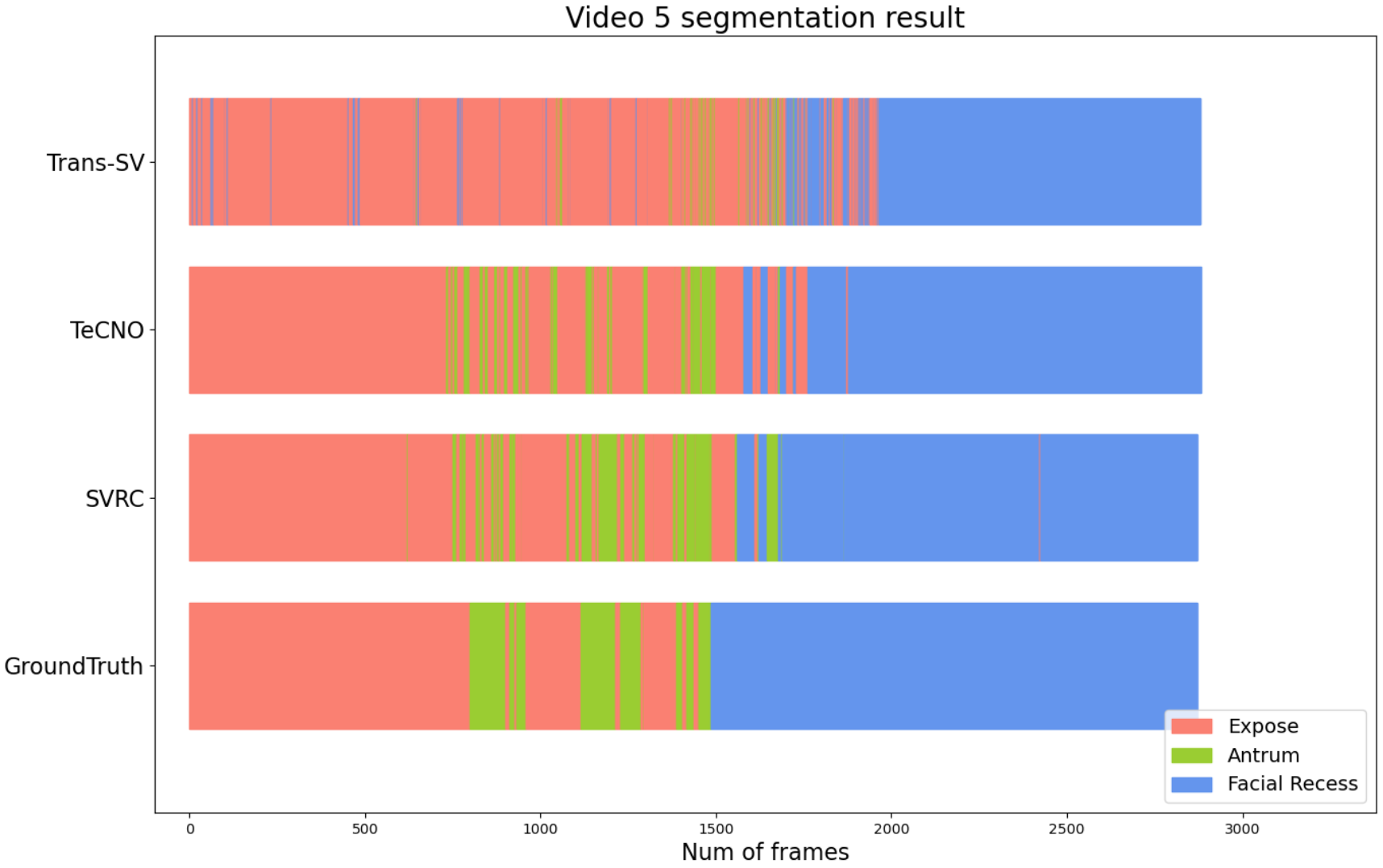
Training (11 videos):

	SVRC (ResNet + LSTM)	TeCNO (ResNet+MSTCN)	Trans-SV (ResNet+MSTCN+Transformer)
Avg. Accuracy	**0.993**	0.979	0.832
Avg. Precision	**0.990**	0.972	0.792
Avg. Recall	0.993	**0.980**	0.812

Validation(2 videos):

	SVRC (ResNet + LSTM)	TeCNO (ResNet+MSTCN)	Trans-SV (ResNet+MSTCN+Transformer)
Avg. Accuracy	**0.88**	0.87	0.77
Avg. Precision	**0.76**	0.65	0.591
Avg. Recall	**0.76**	0.69	0.612

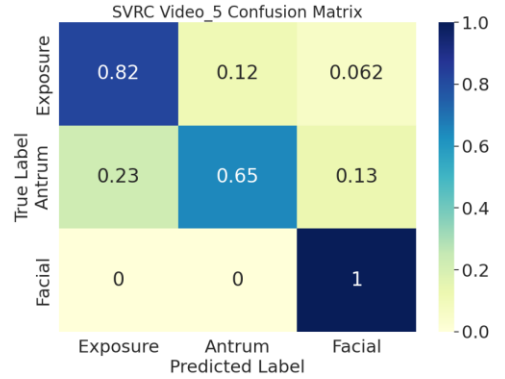
Current Progress: Existing Methods Benchmarking -- Validation Video 5



Current Progress: Existing Methods Benchmarking -- Validation Video 5

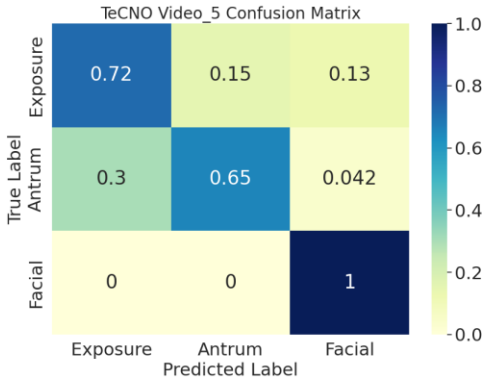
	Expose	Open Antrum	Facial Recess
Accuracy	0.90	0.90	0.96
Precision	0.96	0.61	0.91
Recall	0.82	0.65	1.0

SVRC(ResNet + LSTM) Video 5 results



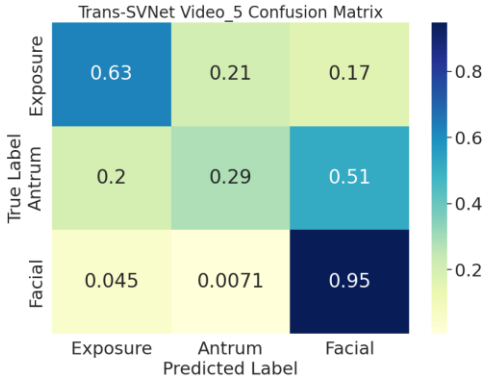
	Expose	Open Antrum	Facial Recess
Accuracy	0.83	0.89	0.93
Precision	0.94	0.41	0.86
Recall	0.72	0.65	1.0

TeCNO Video 5 results

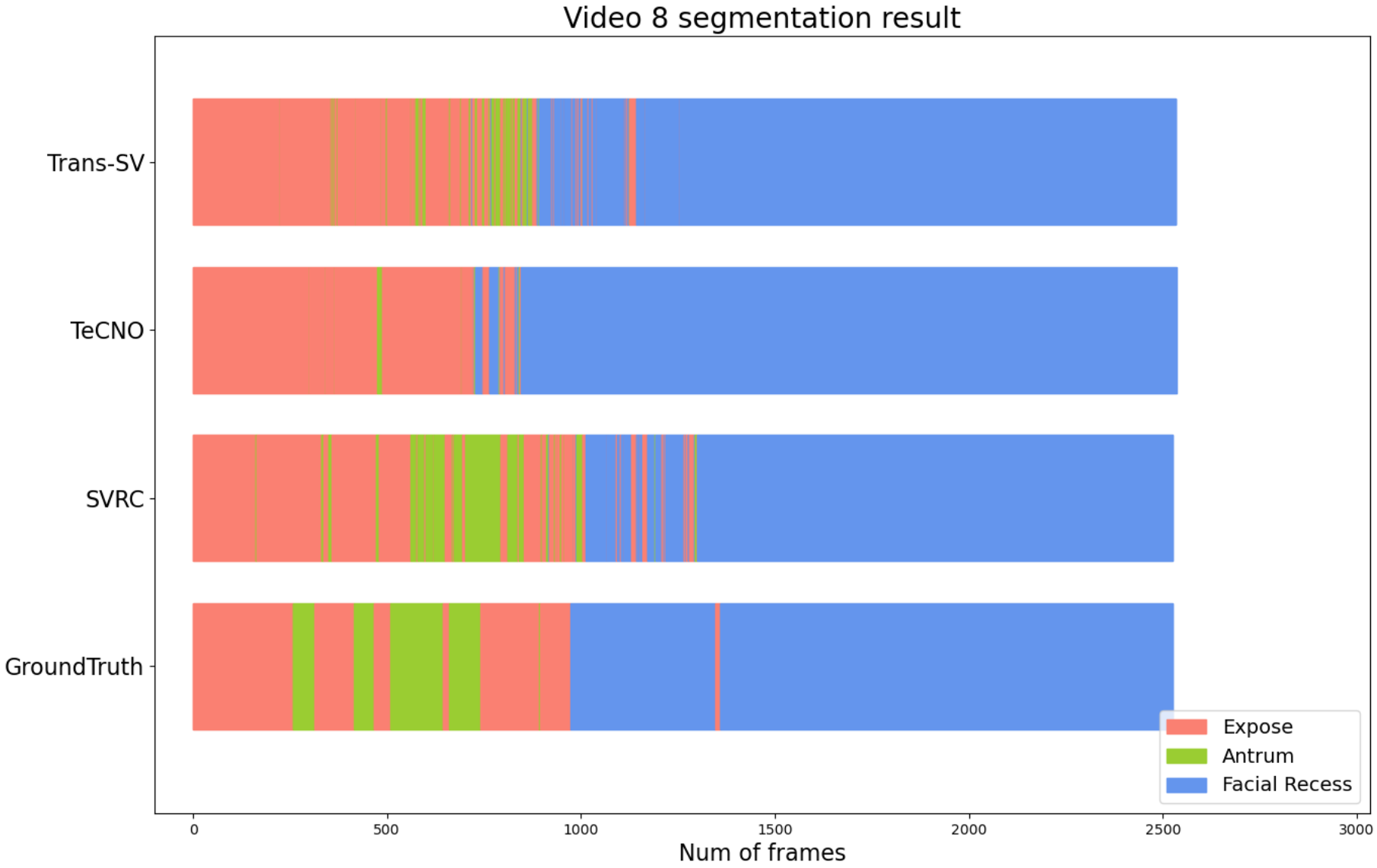


	Expose	Open Antrum	Facial Recess
Accuracy	0.76	0.86	0.87
Precision	0.94	0.05	0.76
Recall	0.63	0.29	0.95

Trans-SVNet Video 5 results



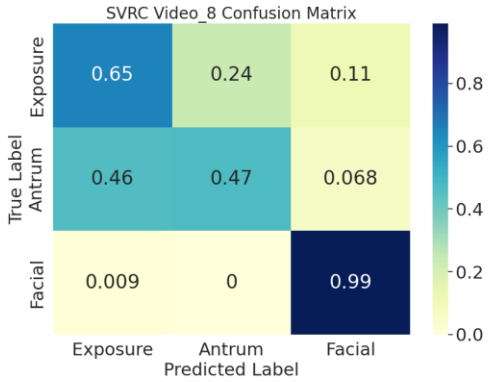
Current Progress: Existing Methods Benchmarking -- Validation Video 8



Current Progress: Existing Methods Benchmarking -- Validation Video 8

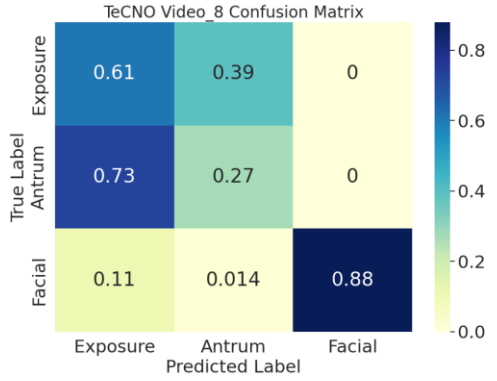
	Expose	Open Antrum	Facial Recess
Accuracy	0.83	0.86	0.95
Precision	0.77	0.43	0.93
Recall	0.65	0.47	0.99

SVRC(ResNet + LSTM) Video 8 results



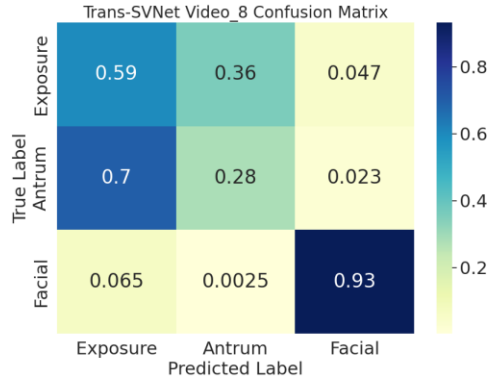
	Expose	Open Antrum	Facial Recess
Accuracy	0.80	0.87	0.91
Precision	0.70	0.018	1.0
Recall	0.61	0.27	0.88

TeCNO Video 8 results



	Expose	Open Antrum	Facial Recess
Accuracy	0.80	0.84	0.94
Precision	0.71	0.11	0.97
Recall	0.59	0.28	0.93

Trans-SVNet Video 8 results



Current Progress: Existing Methods Benchmarking -- Analysis

- Low precision and Recall on "Open Antrum" label for all methods.
- Transformer based method not converging well.
- The validation predictions are unsmooth for all methods. -> Work on transition phase.
- The most straightforward model (SVRC) turns out to be more generalized, and has the best validation result in our database.
- End-to-end models are easier to be transferred to our task, as less "fine-tuning" is needed.

Current Progress: DL Framework for Mastoid Dataset

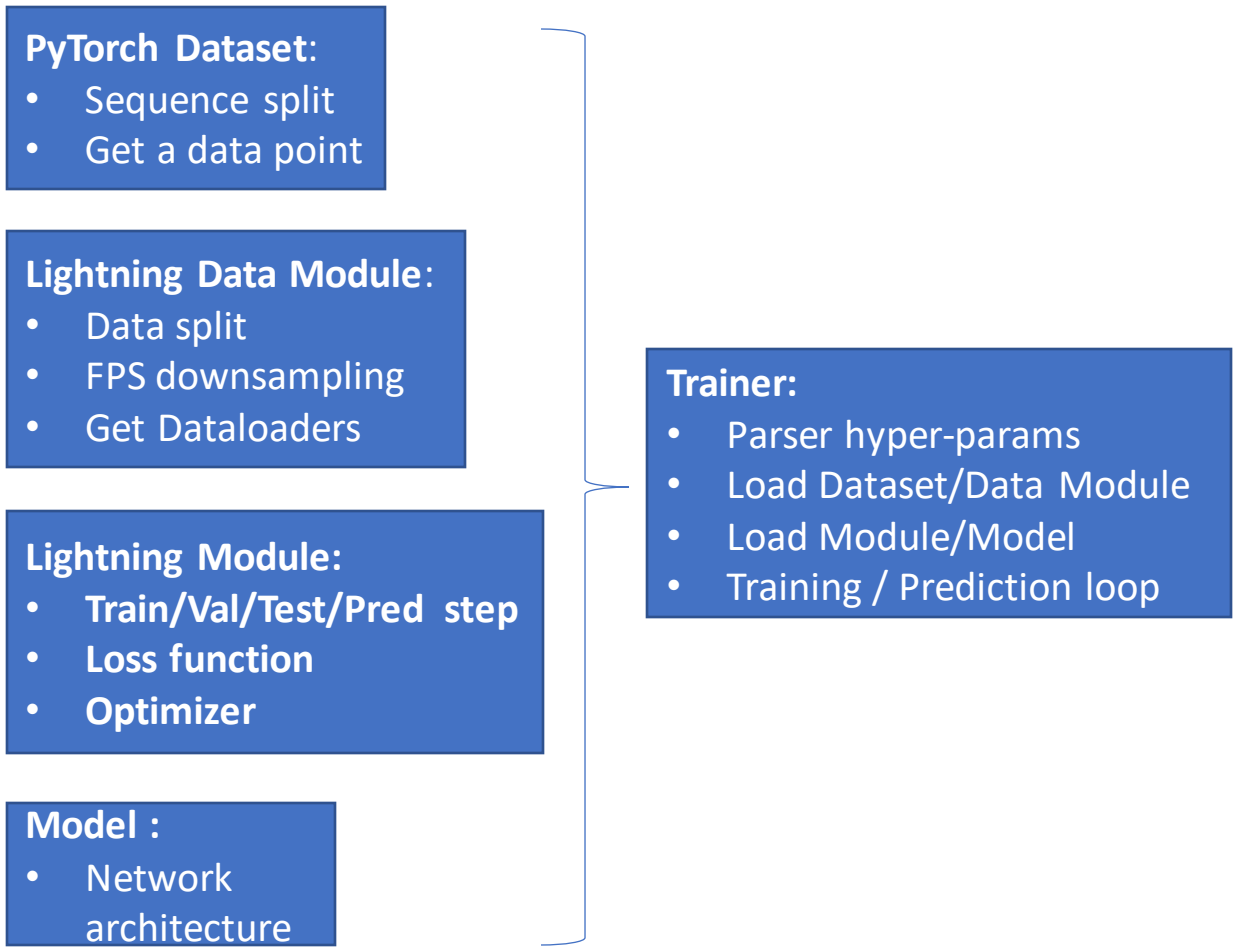
- **Overview**:

  A **Python** project based on **PyTorch Lightning,** implementing **basic data management** and **training process** for **DL methods** on Mastoid dataset

- **Motivations**:
  - Provide an easy way to **implement new** and **benchmark existing** DL methods on Mastoid dataset
  - Can be easily deployed/tested by others
- **Features**:
  - Easy environment setup with docker
  - Easy hyperparameters setup with config file
  - Dataset/Datamodule base class managing data split, downsampling, sequence split
  - Trainer base class managing training/testing/predicting logics
  - Detailed documentations

Current Progress: DL Framework for Mastoid Dataset

Framework Structure:



Reusable code for all methods:

- Dataset(base class)
- Data Module
- Trainer (base class

Code needed for new methods:

- Dataset-specific __getitem(index)__
- Model architecture and forward step
- Optional: prediction logics

Current Progress: DL Framework for Mastoid Dataset

- **Documentation Sample:**

```python
class MastoidDatasetBase(Dataset):
    """ Mastoidectomy Surgical Phase Segmentation Dataset
        base class
    """

    def __init__(self, hparams, df: pd.DataFrame, seq_length: int,
                 video_indexes: List[int],
                 transform: Optional[Compose] = None,) -> None:
        """ MastoidDatasetBase Constroctor

        Args:
            hparams (_type_): hyperparameters and setting read from config file
            df (pd.DataFrame): DataFrame containing metadata for all videos.
                Required columns:
                    1. data file path
                    2. surgical phase label
                    3. video index to which data file corresponds
            seq_length (int): sequence length.
                Each data point is a subsequence of data (i.e raw image, spatial/temporal feature)
                from same video.
                Use seq_length = 1 for per-frame data.
                Use seq_length = -1 if each data point is a video. Lengths of videos
                is stored in self.video_lengths.
            video_indexes (List[int]): list of video indexes.
            transform (Optional[Compose], optional): data transform. Defaults to None.
        """
```

Challenges & Future Work:

1. Limited Dataset.
2. Many transitions within a short period.
3. Effective spatial embedding
4. Impairing benchmarking network over mastoidectomy surgery

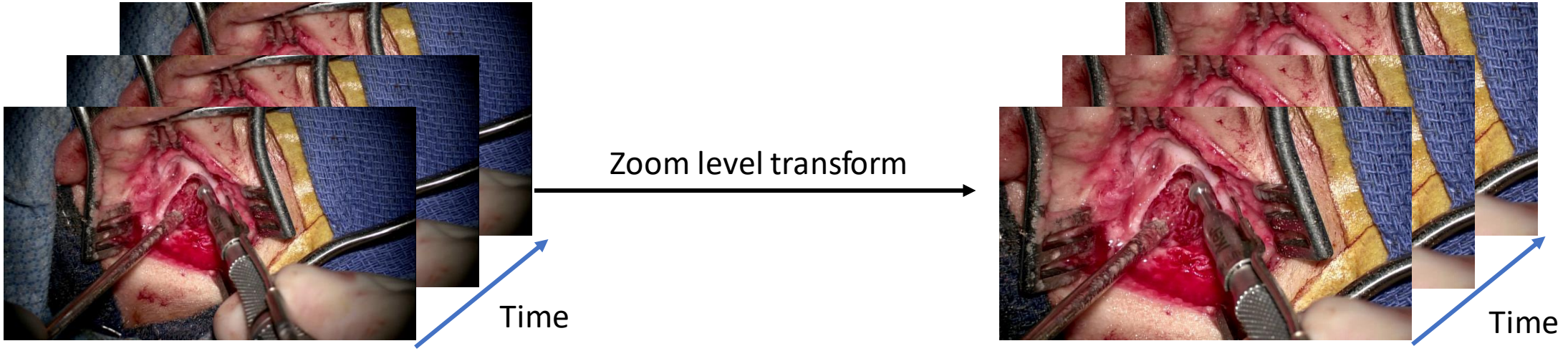
Challenges & Future Work:

Limited Dataset

1. Currently, we have 14 videos captured from different cameras and surgeons. (Different ROI)
2. Different lighting conditions and anatomical appearance.

Proposed Solution:

1. Cross-Validation Training. (Has been implemented manually)
2. Enrichment of our dataset. (Has been resolved by requesting more videos from our mentor)
3. Group data augmentation: Zoom level transform of sequential frame
    - Preserving important clue based on our assumptions



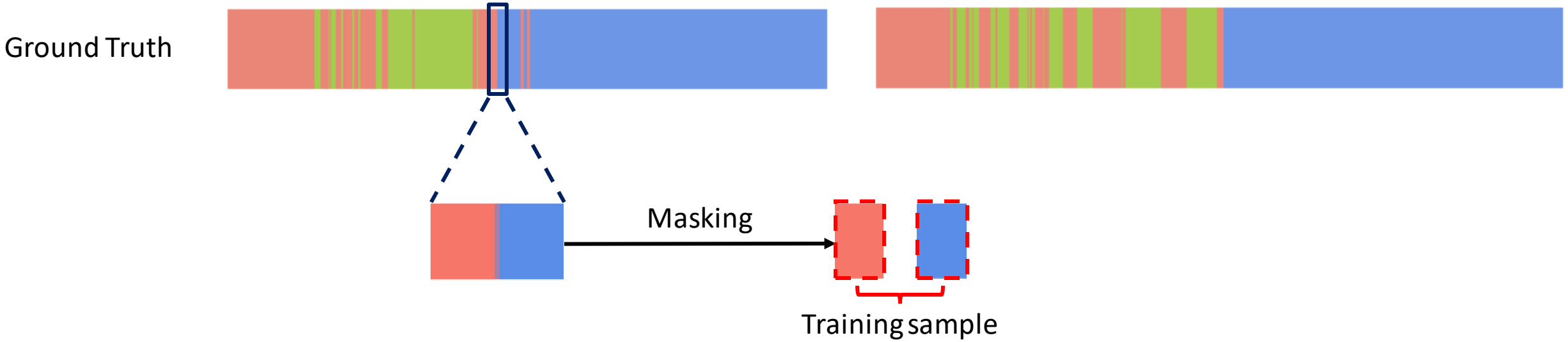
Challenges & Future Work:

Many transitions within a short period

1. Network performance could be undermined by trying to "learn" the transitions.[1]
2. The transition boundaries are fuzzy.

Proposed Solution:

1. Masking the transition boundaries during training:
   - Compensate the fuzzy effect.
   - Drive the network to focus on accurate time stamp.



[1]Wang, Zixun, et al. "Less is More: Surgical Phase Recognition from Timestamp Supervision." *arXiv preprint arXiv:2202.08199* (2022).

Challenges & Future Work:

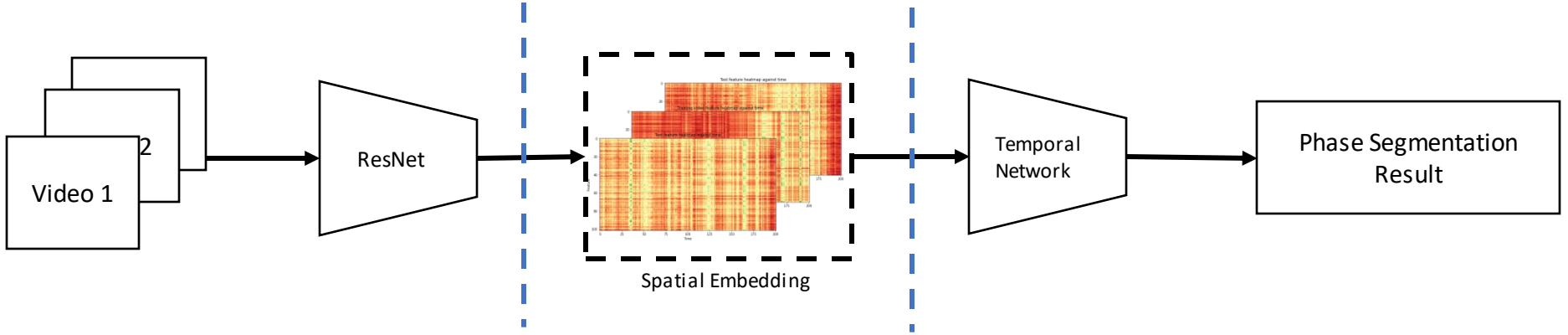
Effective spatial embedding

1. The second stage temporal network performance relies on embedding from the spatial feature extractor.
2. Hard to evaluate the effectiveness of the embedding.(The spatial embedding is uninterpretable)

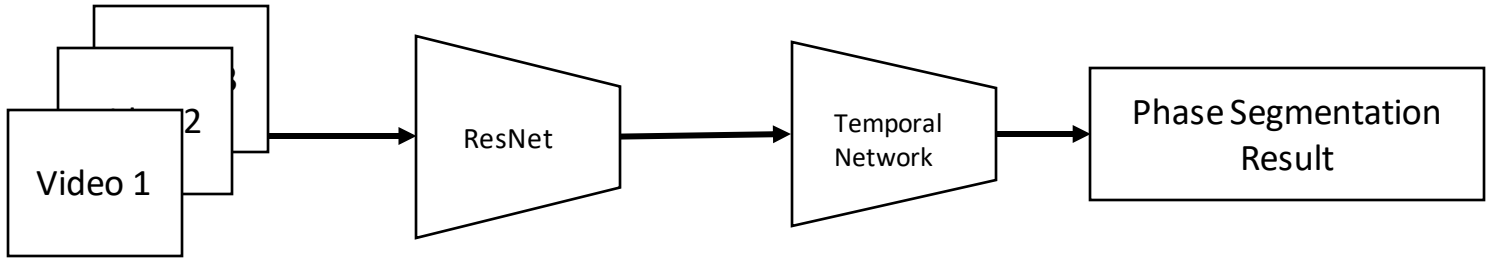
Proposed Solution:

1. End-to-End training

Two-stage Architecture:



End-to-End:



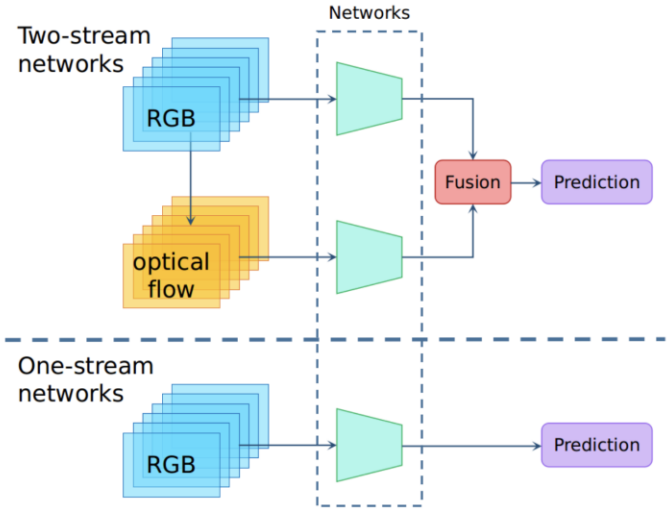
Challenges & Future Work:

Impairing benchmarking network over mastoidectomy surgery

1. The Cholec methods has loss function on tool presence.
   - Assumptions of benchmarking methods over Cholec are related to tool presences.
   - For mastoidectomy surgery → Zoom level change and Tool movement
2. Spatial embedding are the latent features (uninterpretable)

Future work:

1. Specific loss function design based on the proposed assumptions.
2. Interpretable network design (e.g. Optical Flow[1][2])



[1] Wang, Chenhao, et al. "Rgb stream is enough for temporal action detection." *arXiv preprint arXiv:2107.04362* (2021).
[2] Lin, Chuming, et al. "Learning salient boundary feature for anchor-free temporal action localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Milestone Status:

1. Milestone name: Proposal and Plan
   - Planned Date: 02/10
   - Expected Date: 02/10
   - Status: 100%
2. Milestone name: Sample Dataset
   - Planned Date: 02/20
   - Expected Date: 02/20
   - Status: 100%
3. Milestone name: Fully Annotated Dataset
   - Planned Date: 03/17
   - Expected Date: 04/17
   - Status: 12/15 80%
4. Milestone name: Minimum Deliverables
   - Planned Date: 03/27
   - Expected Date: 03/27
   - Status: 100%
5. Milestone name: Initial Network Design
   - Planned Date: 04/22
   - Expected Date: 04/22
   - Status: 30%
6. Milestone name: Expected Deliverables
   - Planned Date: 04/29
   - Expected Date: 04/29
   - Status: 60%
7. Milestone name: Final Presentation
   - Planned Date: 05/02
   - Expected Date: 05/02
   - Status: 0%

Deliverables:	Updated Version	Status
Minimum	1. Same 2. Same 3. A well-documented benchmarking code base that can be easily reused to test out any model for our dataset.	Completed
Expected	1. Same 2. Redesign a new model that achieved 85% accuracy, 80% recall, and precision.	In progress
Maximum	Same	During the summer

Surgical Phase Recognition using Deep Learning

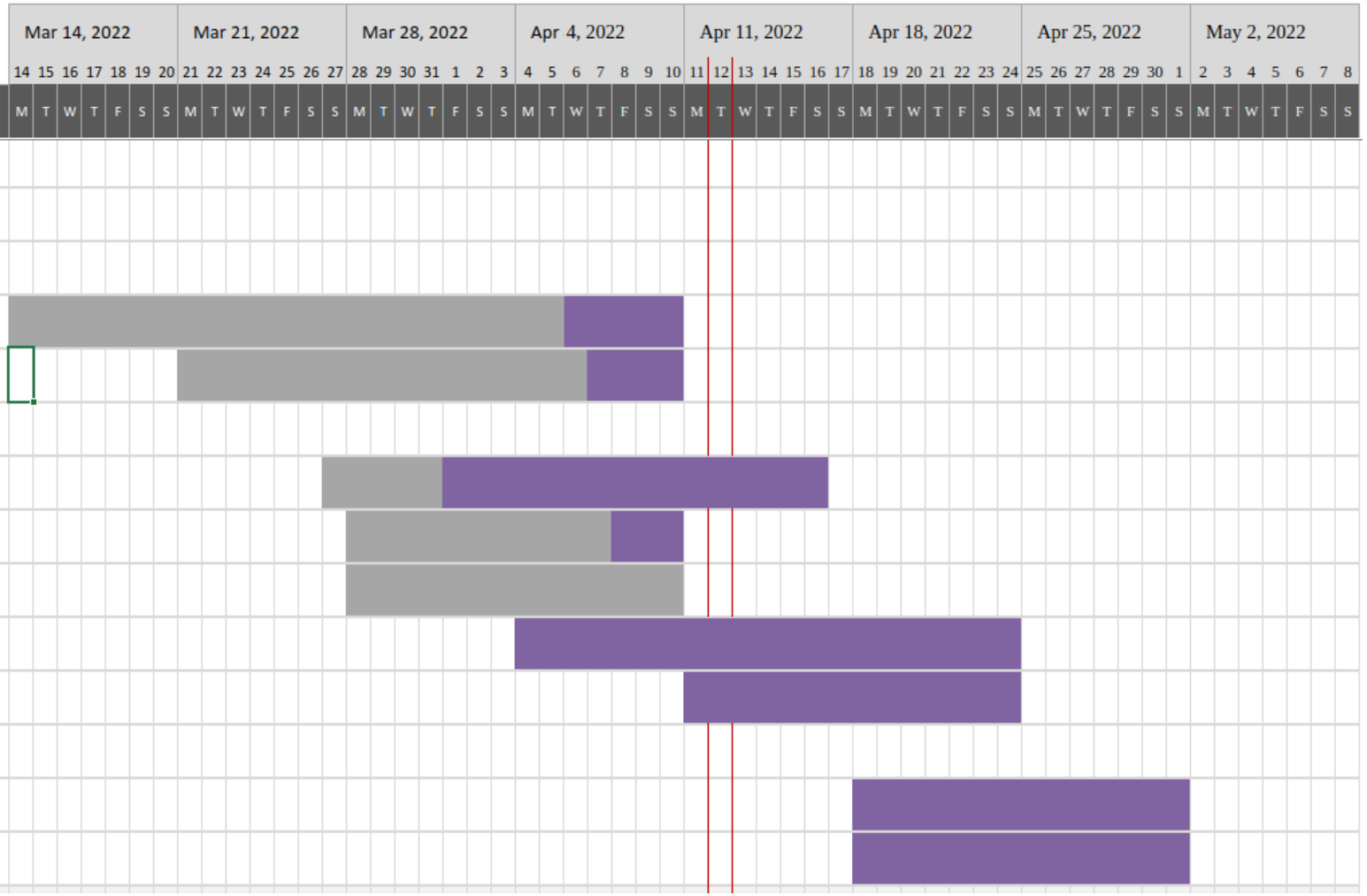
2022 Spring CIS II

Xiaorui Zhang
Wenkai Luo
Xucheng Ma

Project Start: Mon, 1/24/2022

Display Week: 8

TASK	ASSIGNED TO	PROGRESS	START	END
Banchmarking Existing Methods				
Data Pre-processing	1 method p.p.	100%	2/21/22	3/6/22
Data Loader & Loss Function	1 method p.p.	100%	2/28/22	3/13/22
Training	1 method p.p.	85%	3/14/22	4/10/22
Experiments & Analysis	All	85%	3/21/22	4/10/22
Designing and Implementing New DL Method				
Designing Method	TBD	25%	3/27/22	4/16/22
Data Pre-processing	TBD	85%	3/28/22	4/10/22
Data Loader & Loss Function	TBD	100%	3/28/22	4/10/22
Training	TBD	0%	4/4/22	4/24/22
Experiments & Analysis	TBD	0%	4/11/22	4/24/22
Results & Presentations				
Poster	All	0%	4/18/22	5/1/22
Final Report	All	0%	4/18/22	5/1/22



Thank you!

Q & A

References

[1] Y. Jin, Q. Dou, H. Chen, L. Yu, J. Qin, C.-W. Fu, and P.-A. Heng, “Sv-rcnet: Workflow recognition from surgical videos using recurrent convolutional network,” IEEE Transactions on Medical Imaging, vol. 37, no. 5, pp. 1114–1126, 2018
[2] Czempiel, T., Paschali, M., Keicher, M., Simson, W., Feussner, H., Kim, S. T., & Navab, N. (2020, October). Tecno: Surgical phase recognition with multi-stage temporal convolutional networks. In International conference on medical image computing and computer-assisted intervention (pp. 343-352). Springer, Cham.
[3] Xiaojie Gao, Yueming Jin, Yong-Hao Long, Qi Dou, and Pheng-Ann Heng. Trans-svnet: Accurate phase recognition from surgical videos via hybrid embedding aggregation transformer.CoRR, abs/2103.09712, 2021.
[4] Wang, Zixun, et al. "Less is More: Surgical Phase Recognition from Timestamp Supervision." arXiv preprint arXiv:2202.08199 (2022).
[5] Wang, Chenhao, et al. "Rgb stream is enough for temporal action detection." arXiv preprint arXiv:2107.04362 (2021).
[6] Lin, Chuming, et al. "Learning salient boundary feature for anchor-free temporal action localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.