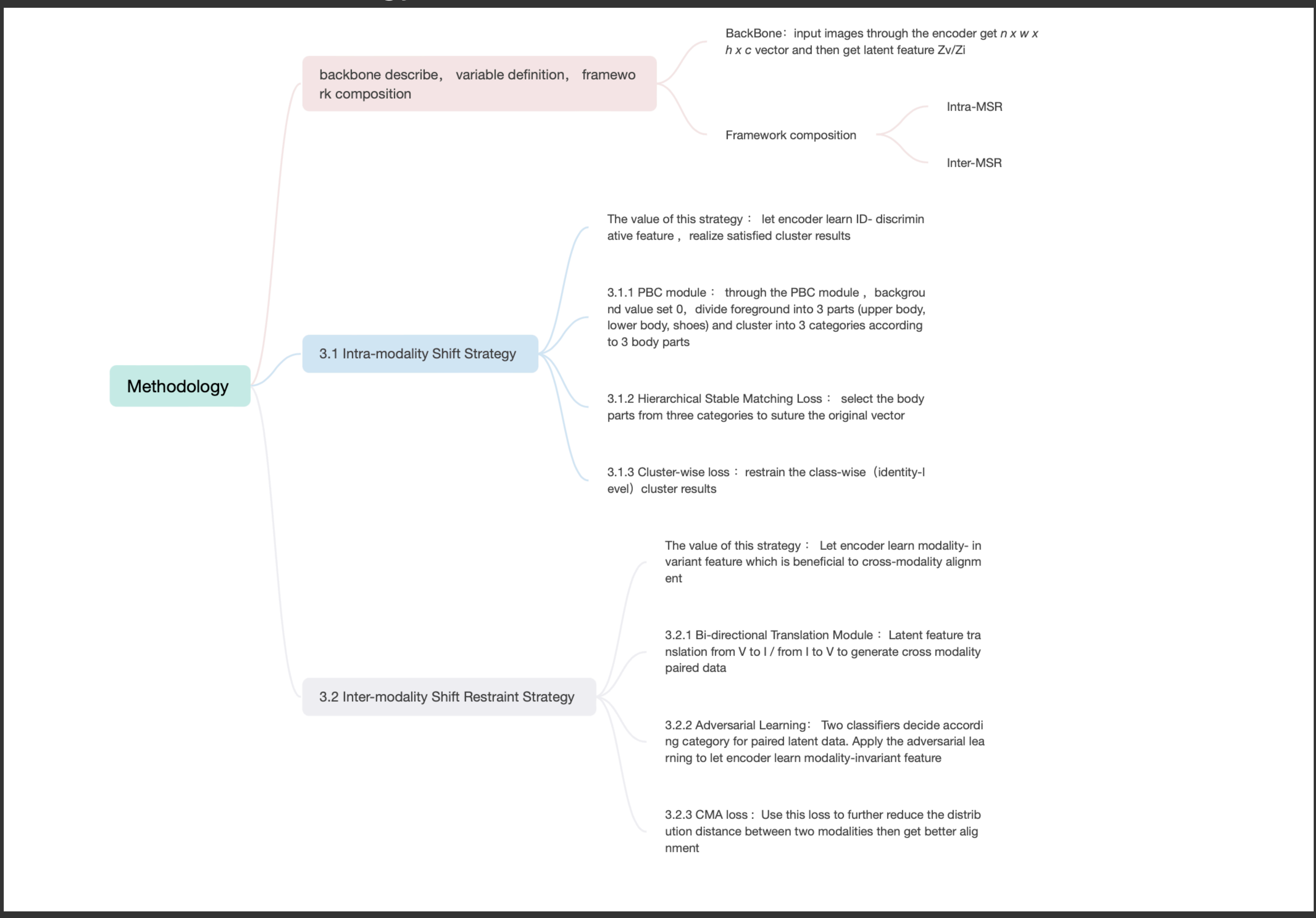
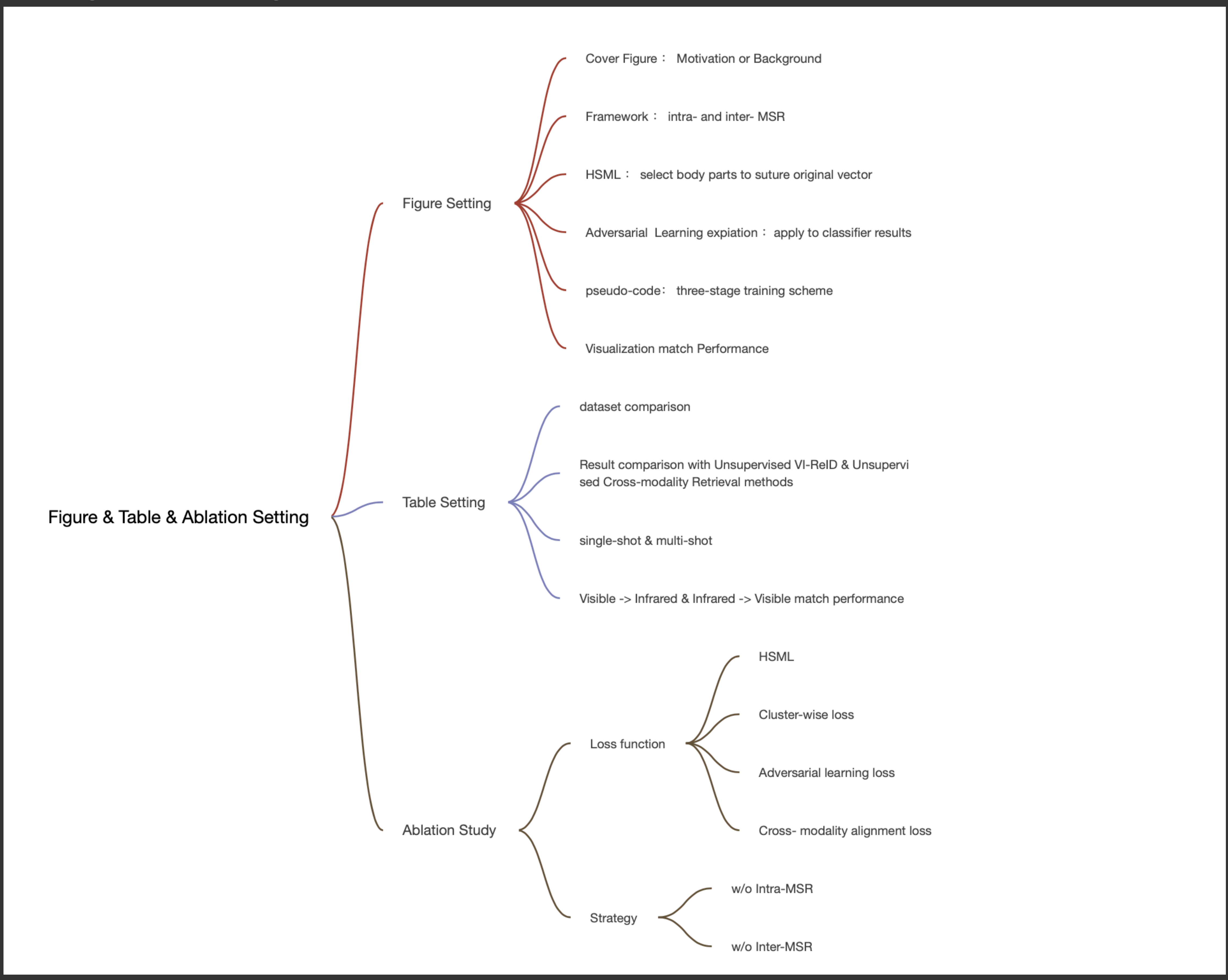
Completed works in this week：

- ☑ Introduction draft
- ☑ Methodology draft
- ☑ Dataset description & Evaluation protocol
- ☑ Experiment Coding — PBC & CMA

Flowchart of Methodology



Arrangement of Figures & Tables



Thoughts on writing the paper:

1. H2H and OTLA are not completely unsupervised methods. H2H directly uses the visible ReID model as a pre-trained model, and OTLA uses well-annotated visible data to assign pseudo-label to unannotated visible data using the UDA method. So our work can be called the first work to realize the fully unsupervised method for VI-ReID.
2. Only using CMA loss is not sufficient, and it may also cause the same identity to not be well aligned in different modalities. But because of the paired data we generate, my solution is:

lead to misalignment due to the confusion of decision boundaries [37, 18]. Therefore, we use multiple alignment losses instead of simple CMA loss. For the visible branch, since we have generated the same identity's visible latent feature in the infrared branch (Z_I^V), we further reduce the distance between two same-modality but different-domain features by applying the distance-of-distance[9] loss to Z_V and Z_I^V. This step is also applied to the infrared branch. Then, we apply the CMA loss to align Z_V and Z_I, Z_V^I and Z_I^V simultaneously, and enforce Lipschitz continuity between them. This further forces the encoder and generator to learn modality-invariant features. Therefore, the final CMA loss is:

$$\mathcal{L}_{\text{TotalCMA}} = \alpha \left(\mathcal{L}_{\text{DD}} Z_V Z_I^V + \mathcal{L}_{\text{DD}} Z_I Z_V^I \right) + \mathcal{L}_{\text{CMA}} Z_V Z_I + \beta \mathcal{L}_{\text{CMA}} Z_V^I Z_I^V \qquad (5)$$

This can make different decision boundaries easier to distinguish and increase the performance of Cross-modality Alignment