- So far, we have discarded 3D object detection and only remain **3D Semantic Segmentation** (lidar segmentation).

- Domain adaptation is still preserved; We train our model with both labeled source data and unlabeled target data and test our model on target_test.

- We adopt nearly the same architecture as xMUDA, except for the head part, where xMUDA introduces "Dual Head" but we only use "Vinilla Fusion". The backbones are the same: UNet for 2D images and SparseConvNet for 3D lidar points.

- The segmentation labels we use are released after xMuda. So we have rerun xMUDA with the new labels. See the results.

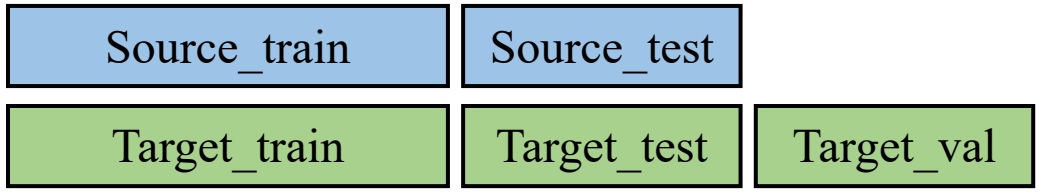
- We have run baseline for our "Vanilla Fusion". Though seemingly strange, but it has outperformed xMUDA. See the results.

- We have tried to add contrastive losses on both source and target training data. But it doesn't give us a better result on target_test. This corresponds to config meta "contrast_usa". See the results.

- So at this point we return to contrastive loss on a single domain to check its effectiveness on source_test. This corresponds to config meta "src_ctr_usa". See the results.

Nuscenes Dataset Domain Splits

- Two Source-Target pairs
  - (USA, Singapore) & (Day, Night)
- Two splits on source
  - Source_train & Source_test
- Three splits on target
  - Target_train, Target_test & Target_val
- All of the experiments (results shown in tables) up to now are carried on these splits!



	source		target		
Split	train	test	train	val	test
Day - Night	24,745	5,417	2,779	606	602
Boston - Singapore	15,695	3,090	9,665	2,770	2,929
A2D2 - SemanticKITTI	27,695	942	18,029	1,101	4,071

Table 4: Number of frames for the 3 splits.

Config Metas:

- baseline_usa

  $\boldsymbol{L = L_{seg}}$

- src_ctr_usa

  $\boldsymbol{L = L_{seg} + \lambda * L_{src\_contrast}}$

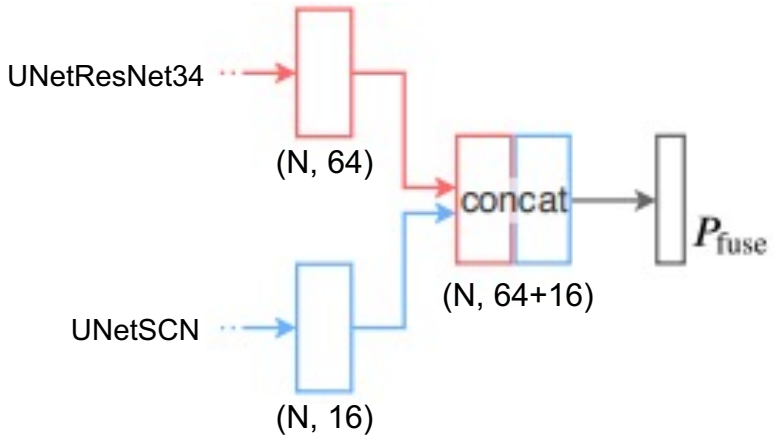
- only_ctr_usa

  $\boldsymbol{L = L_{src\_contrast}}$, or $\boldsymbol{L = L_{src\_contrast} + L_{tgt\_contrast}}$

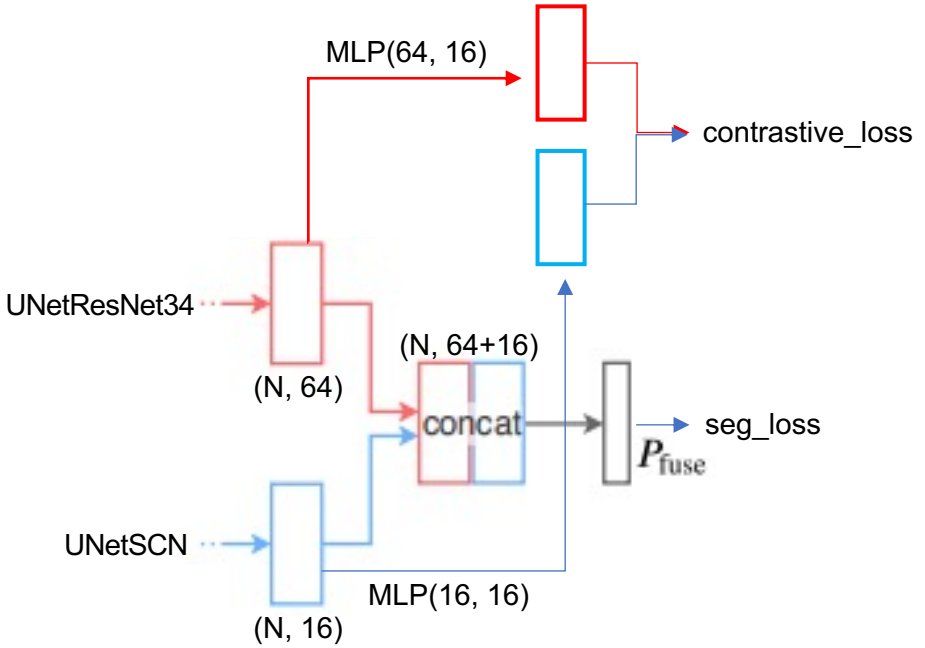
- contrast_usa

  $\boldsymbol{L = L_{seg} + \lambda * (L_{src\_contrast} + L_{tgt\_contrast})}$

Architectures(ours)

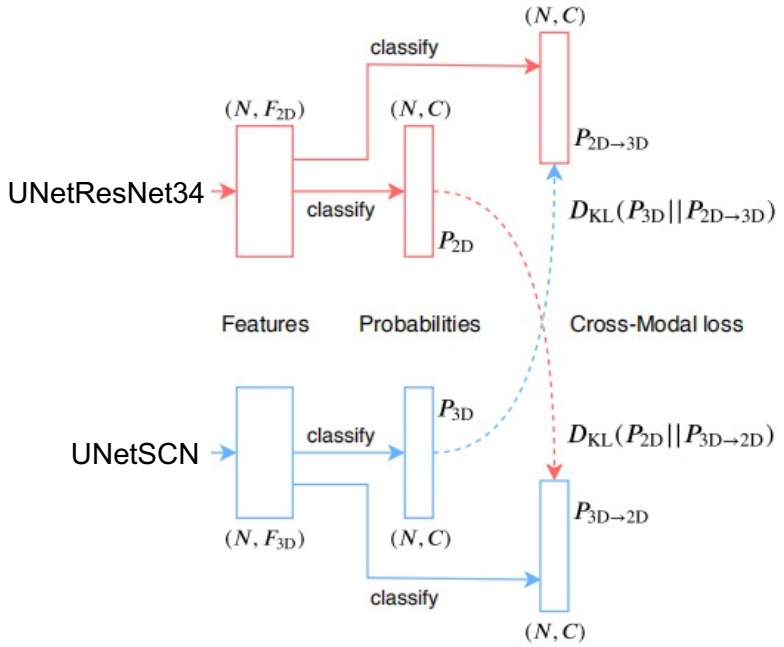


Vanilla Fusion(Baseline)

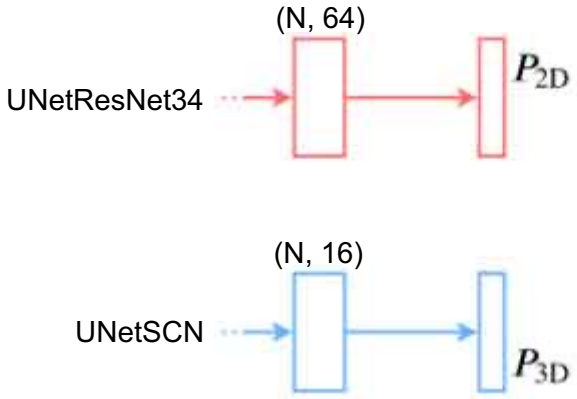


Vanilla Fusion with contrastive loss

Architectures(xMuda)



Dual Head with KL Divergence



Single Head without KL Divergence

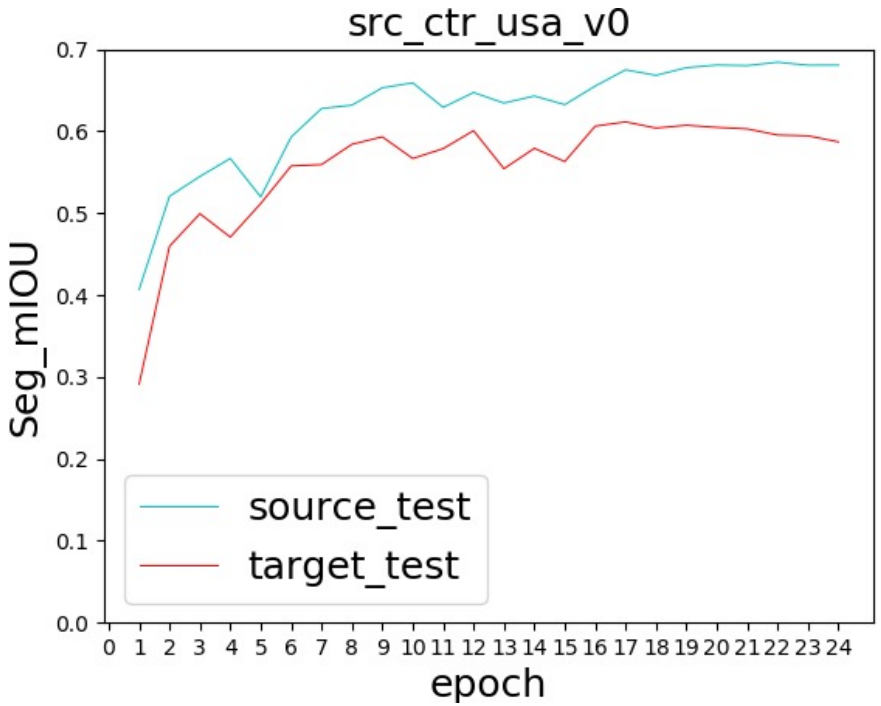
03/28/2021

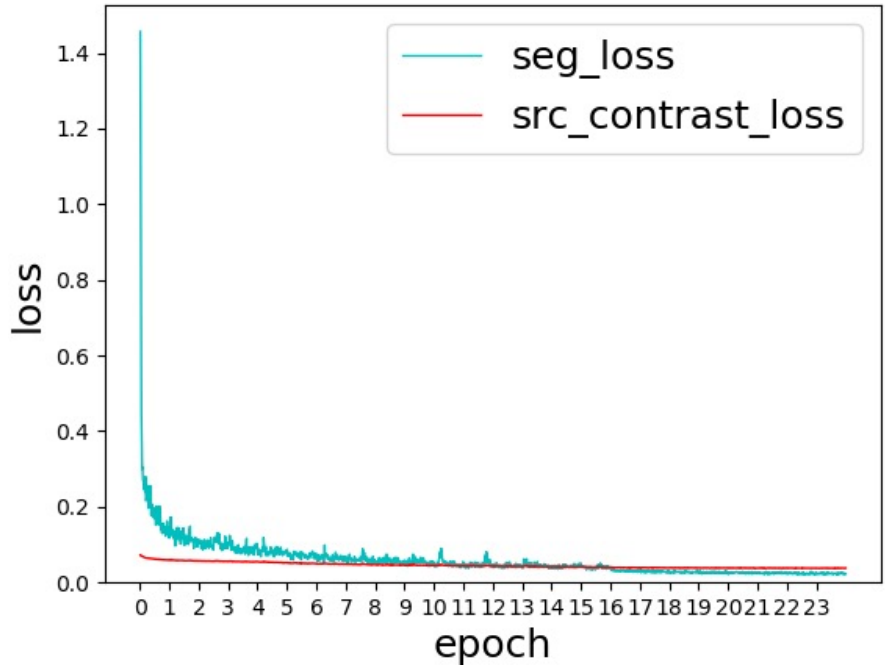
- src_ctr_usa_v3/v4

03/25/2021

- src_ctr_usa_v1/v2
- only_ctr_usa_v0/v1

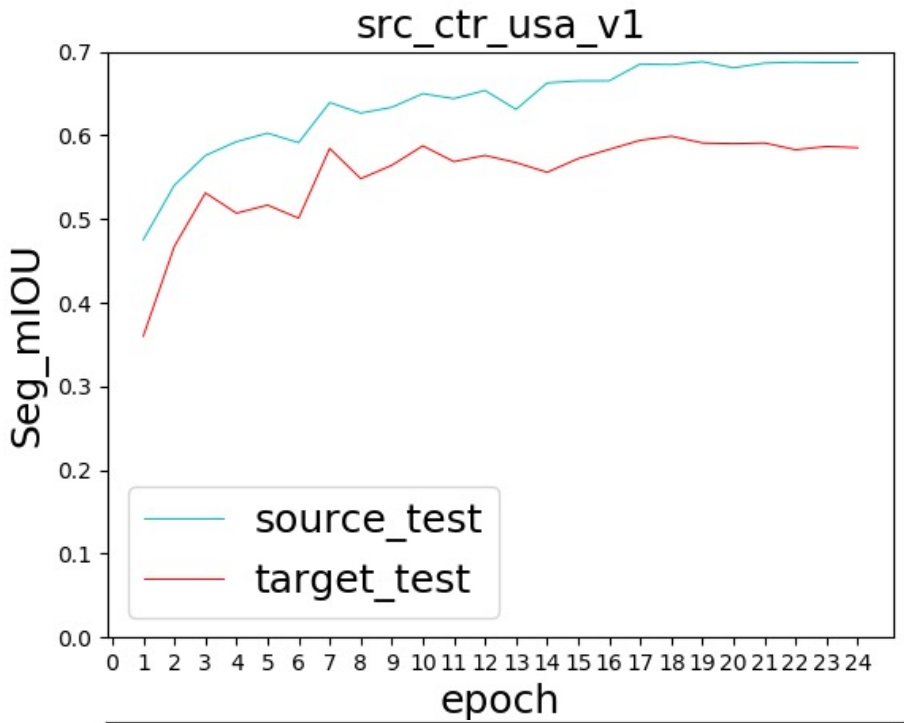
contrastive loss ablation study on single source domain		
Configs	**Source_test (mIOU)**	**Target_test (mIOU)**
src_ctr_usa_v0	68.42	61.14
src_ctr_usa_v1	**68.80**	59.90
src_ctr_usa_v2	65.55	52.04
src_ctr_usa_v3	66.44	54.68
src_ctr_usa_v4	63.66	52.85
baseline_usa_v1	68.01	**63.74**

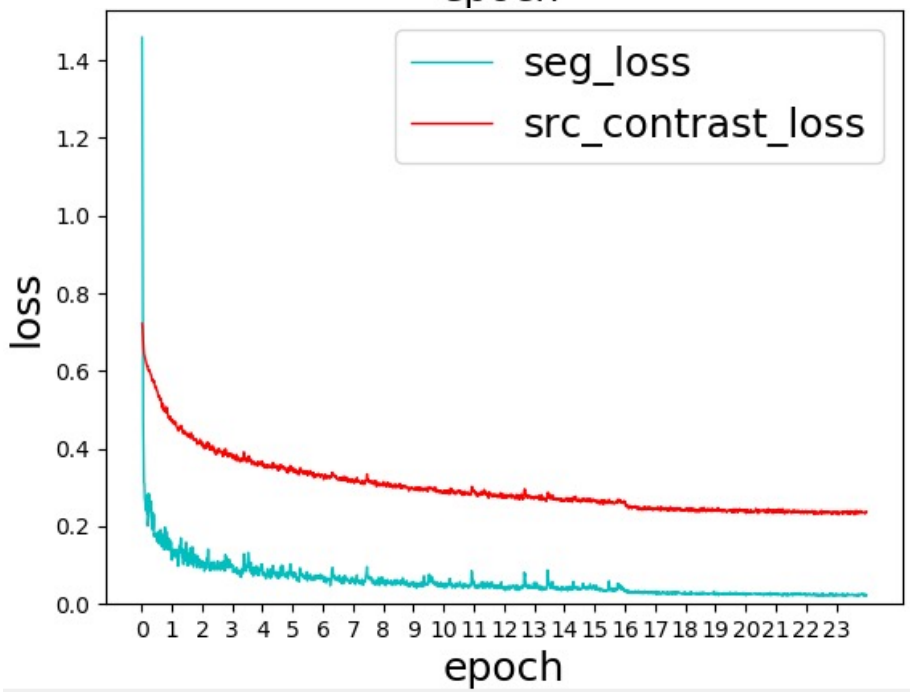




src_ctr_usa_v0:

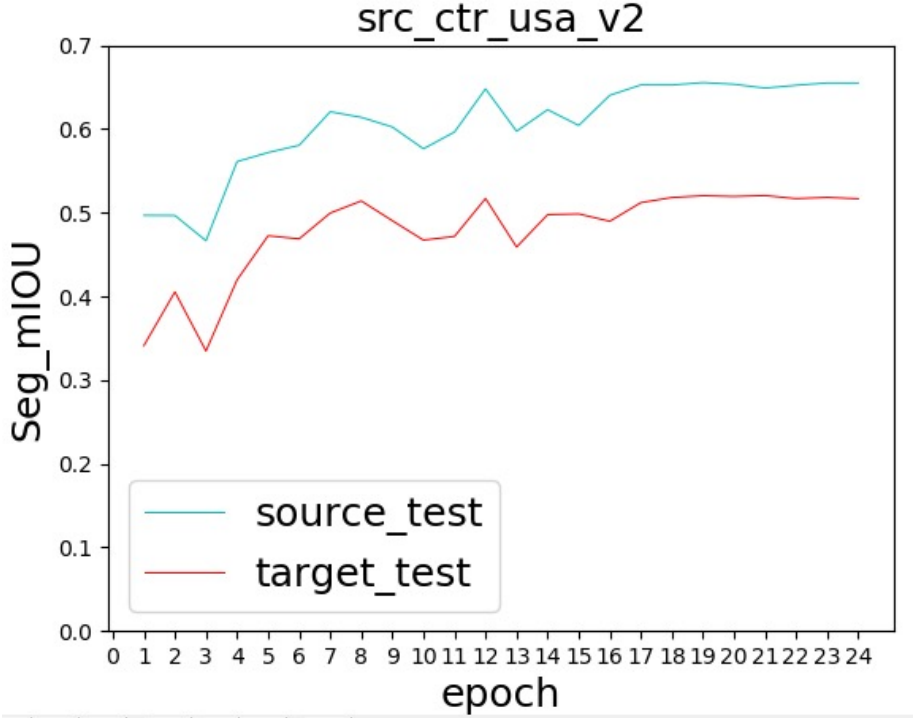
- **$\lambda = 0.01$**
- #pts = 1024, # groups = 1
- batch_size=8
- $L = L_{seg} + \lambda * L_{src\_contrast}$

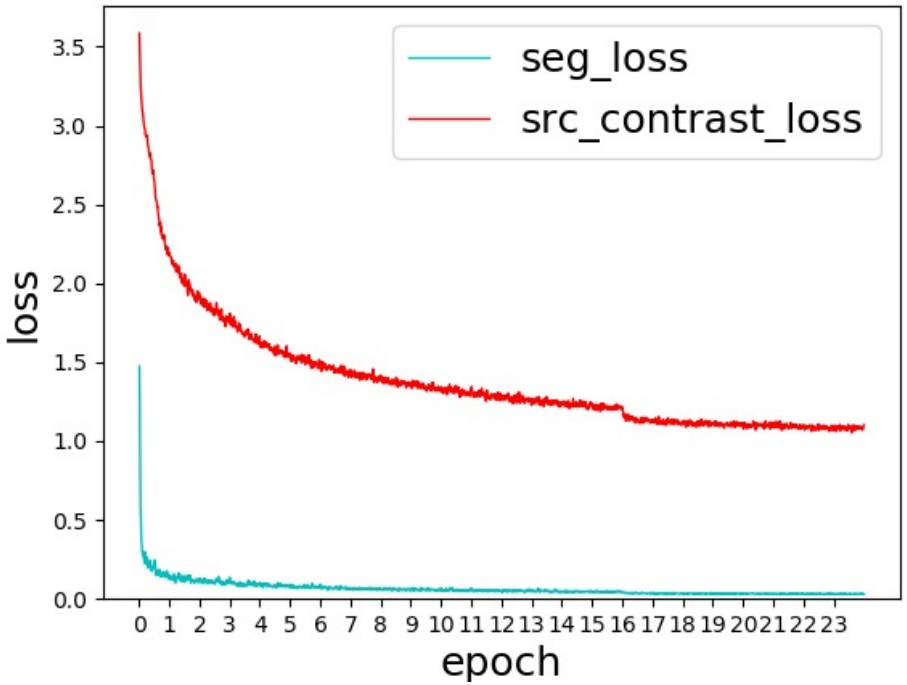




src_ctr_usa_v1:

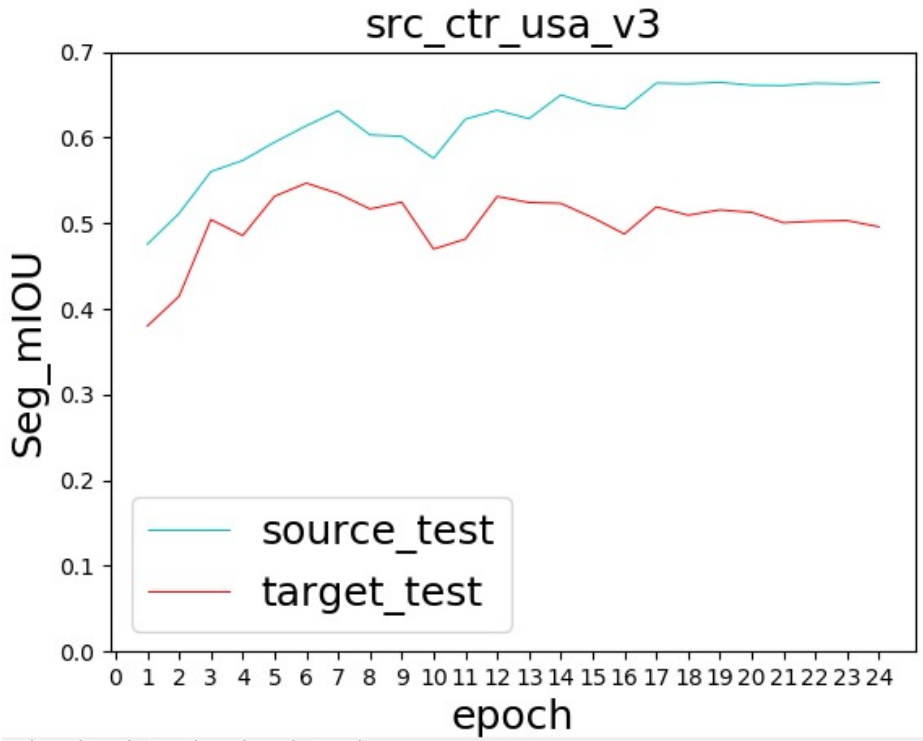
- **$\lambda = 0.1$**
- #pts = 1024, # groups = 1
- batch_size=8
- $L = L_{seg} + \lambda * L_{src\_contrast}$

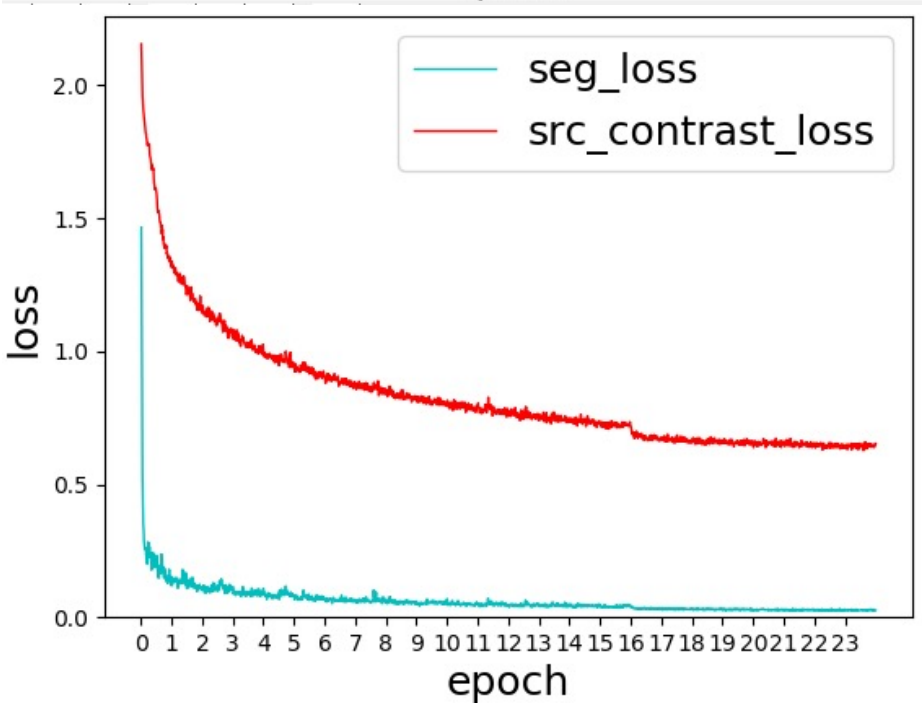




src_ctr_usa_v2:

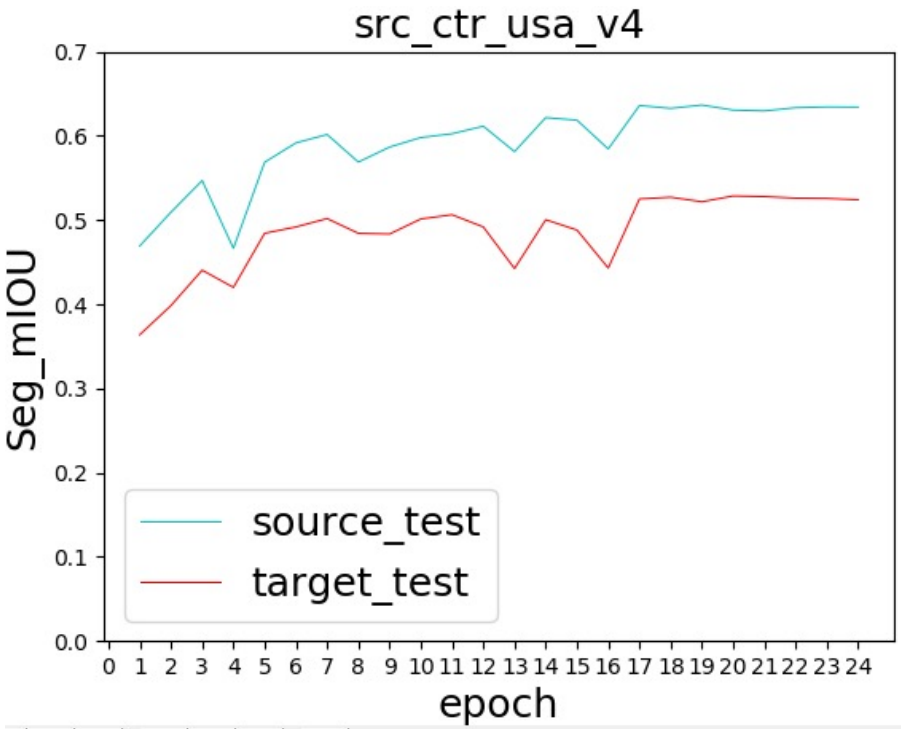
- **$\lambda = 0.5$**
- #pts = 1024, # groups = 1
- batch_size=8
- $L = L_{seg} + \lambda * L_{src\_contrast}$

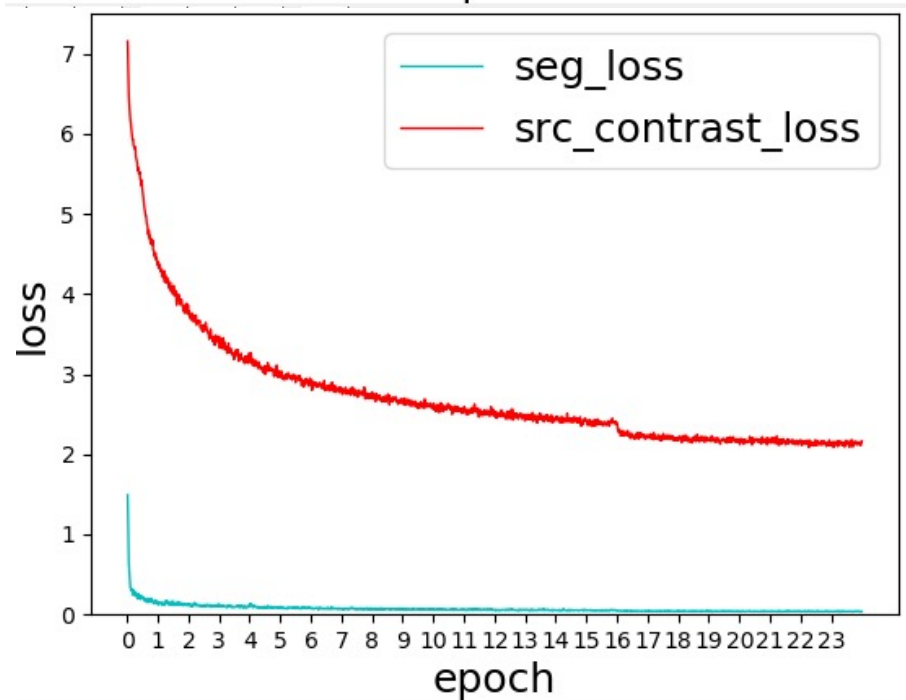




src_ctr_usa_v3:

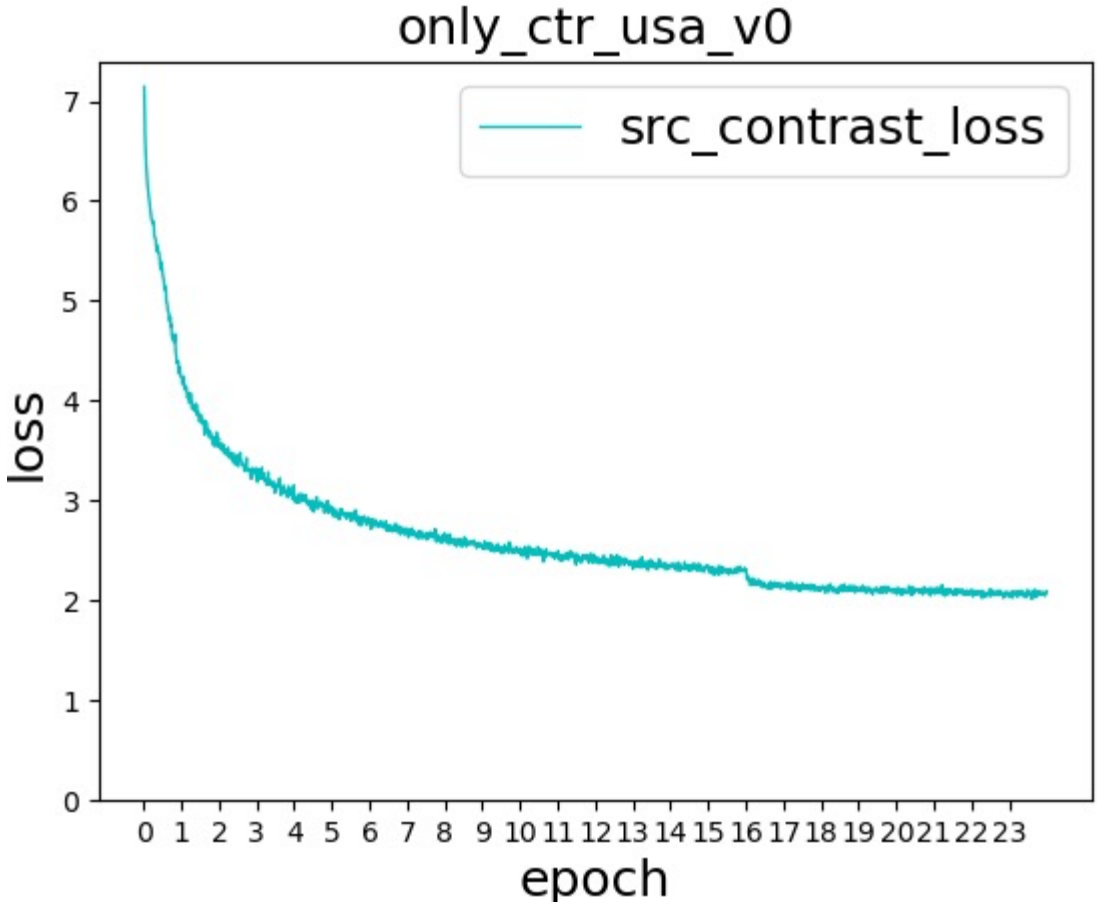
- **$\lambda = 0.3$**
- #pts = 1024, # groups = 1
- batch_size=8
- $L = L_{seg} + \lambda * L_{src\_contrast}$

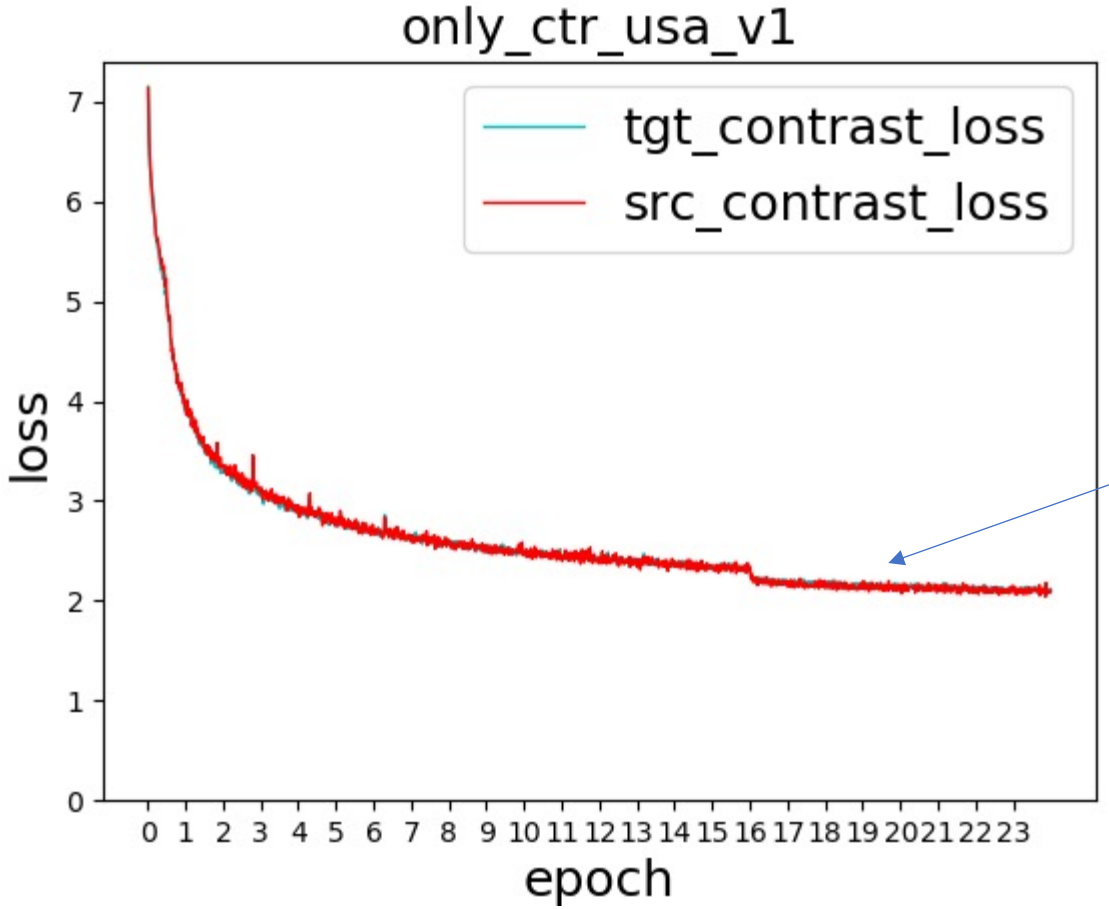




src_ctr_usa_v4:

- **$\lambda = 1.0$**
- #pts = 1024, # groups = 1
- batch_size=8
- $L = L_{seg} + \lambda * L_{src\_contrast}$


only_ctr_usa_v0
src_contrast_loss
loss
epoch
0 1 2 3 4 5 6 7
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23



tgt_contrast_loss & src_contrast_loss nearly **coincide**

03/24/2021

- contrast_usa_v3/v4
- src_contrast_usa_v0

03/20/2021

- baseline_usa_v1(baseline2_usa)
- contrast_usa_v1

03/19/2021

- Baseline_usa_v0(baseline1_usa)
- Contrast_usa_v0

USA/Singapore		
	Config	**Target_test (mIOU)**
xMUDA	dual head + KL_div 2D+3D	62.53
xMUDA baseline	single head 2D+3D	62.10
Vanilla fusion baseline	baseline_usa_v0 (B=4)	61.57
	baseline_usa_v1 (B=8)	**63.74**
Contrastive loss on both source_train & target_train	contrast_usa_v0	52.67
	contrast_usa_v1	61.69
	contrast_usa_v3	< 60.00
	contrast_usa_v4	< 60.00

Hyperparameters to fine-tune:

- λ (0.1, 0.01, 0.005, 0.001)
- Temparature (100, 10, 1, 0.5, 0.1, 0.05)
- #pts in each group (64, 256, 1024, 2048) & #groups in ea
  - $n_{pos}: n_{neg} = 1: (\#pts - 1)$
- Transformation after representation(Linear, Non-Linear, I
- Optimizers(AdamW, Adam, SGD+Momentum)

Meta Config:

contrast_usa:

- batch_size=8
- $L = L_{seg} + \lambda * (L_{src\_contrast} + L_{tgt\_contrast})$

Sub Configs:

contrast_usa_v0:

- $\lambda = 0.1$
- #pts = 1024, # groups = 1

contrast_usa_v1:

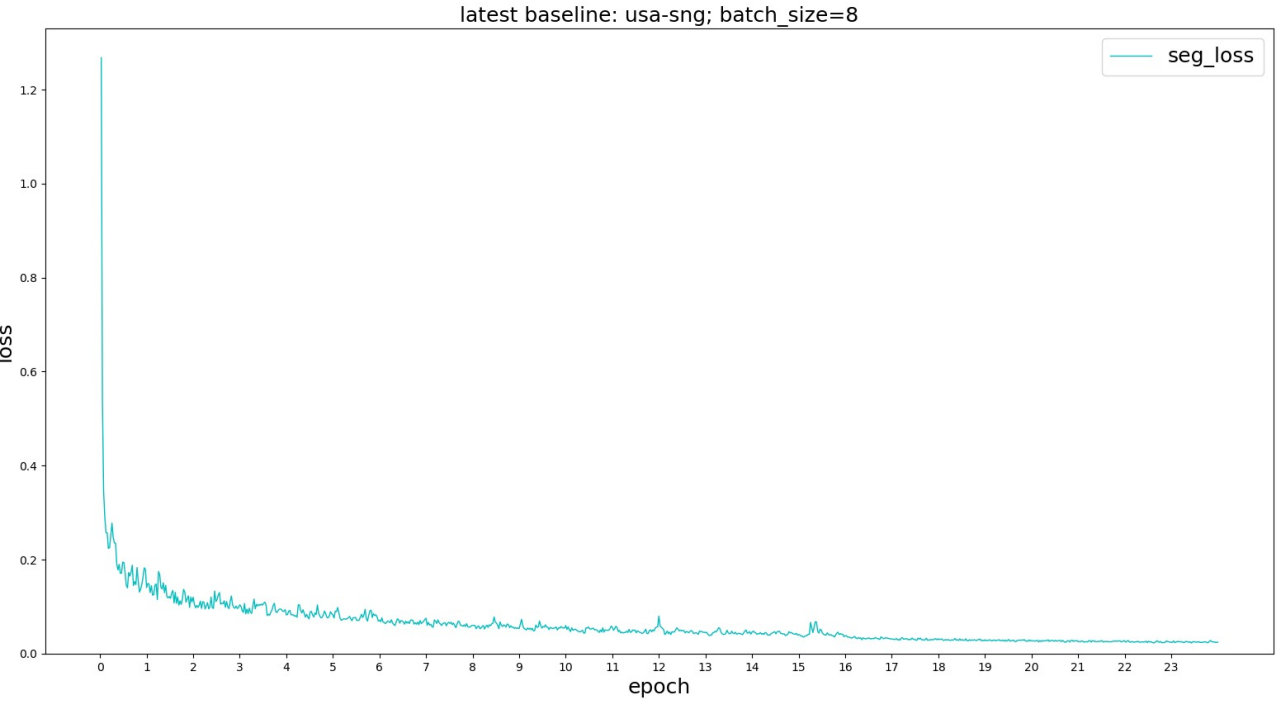
- $\lambda = 0.01$
- #pts = 1024, # groups = 1

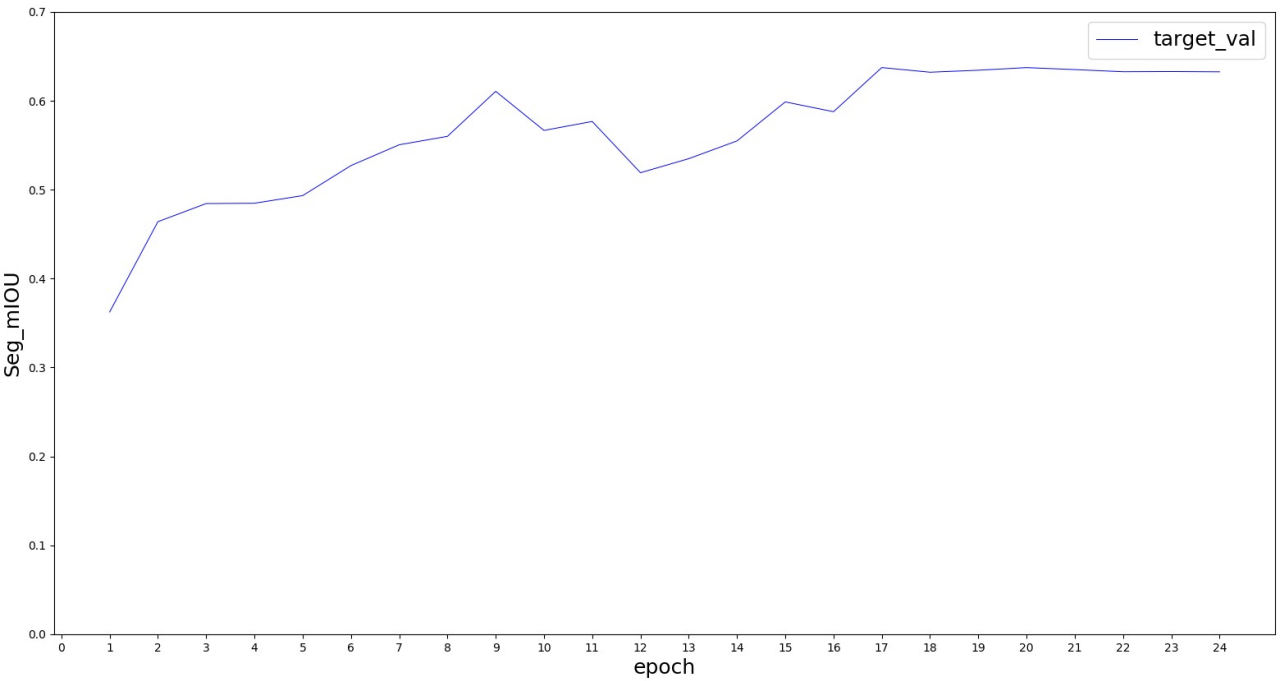
contrast_usa_v3:

- $\lambda = 0.01$
- #pts = 128, # groups = 8

contrast_usa_v4:

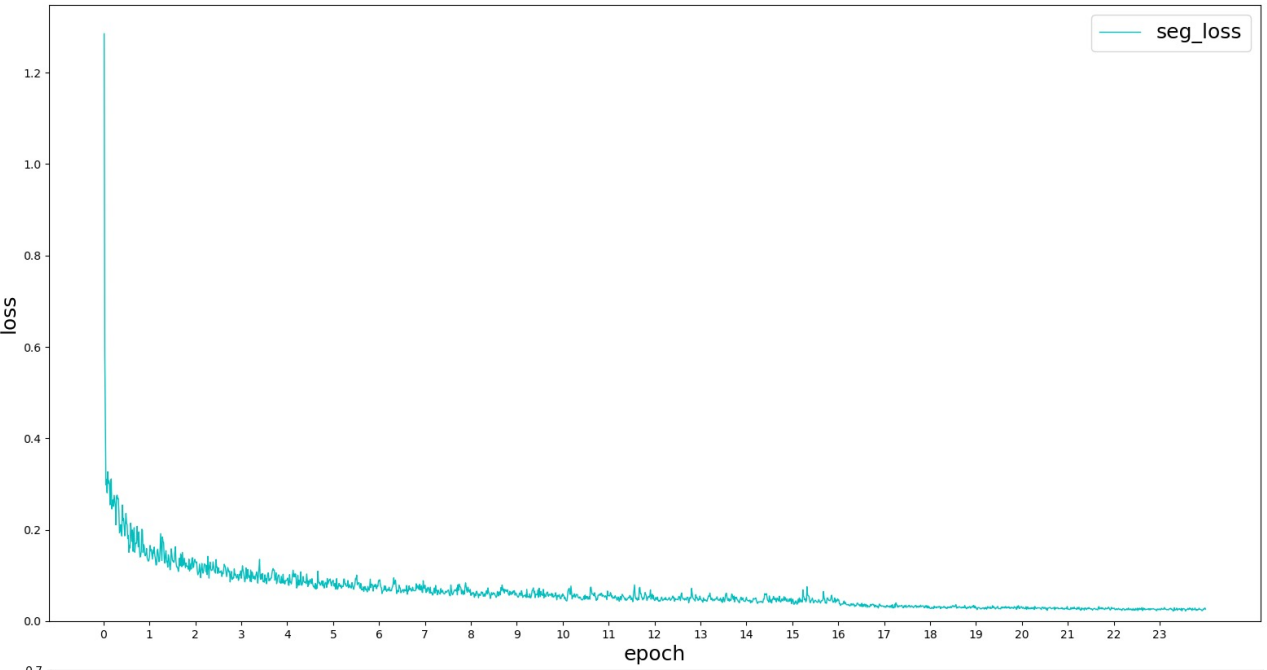
- $\lambda = 0.01$
- #pts = 2048, # groups = 1

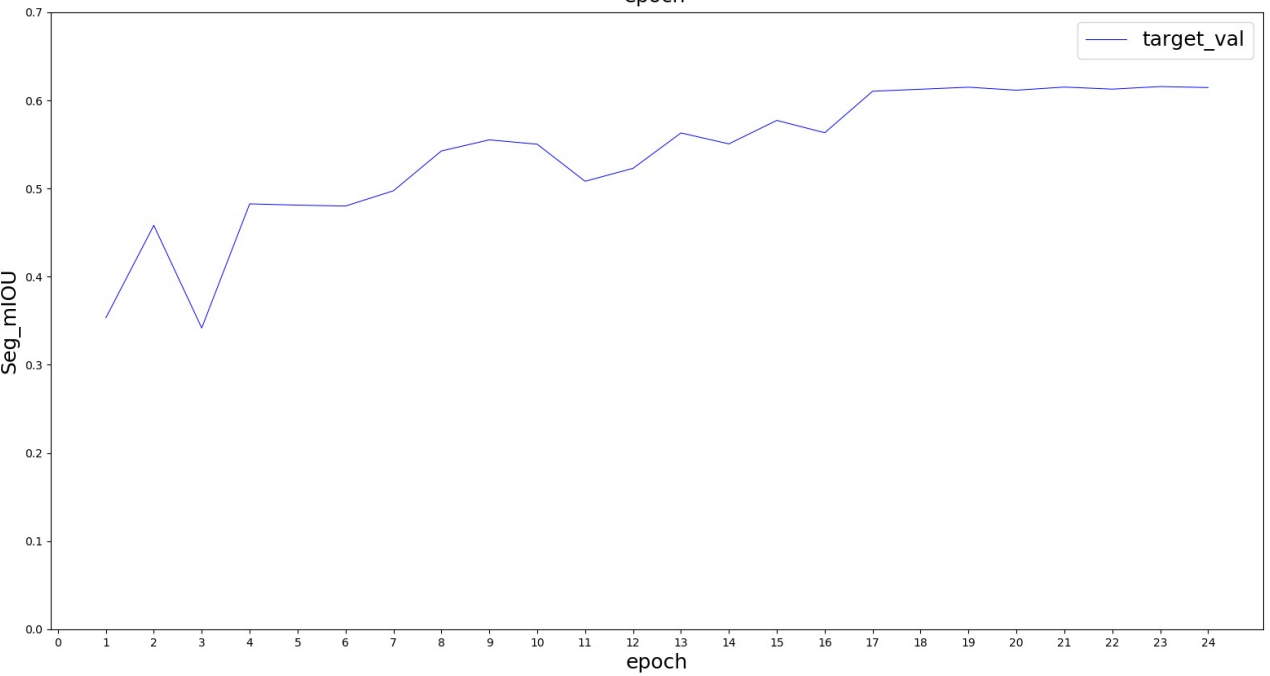


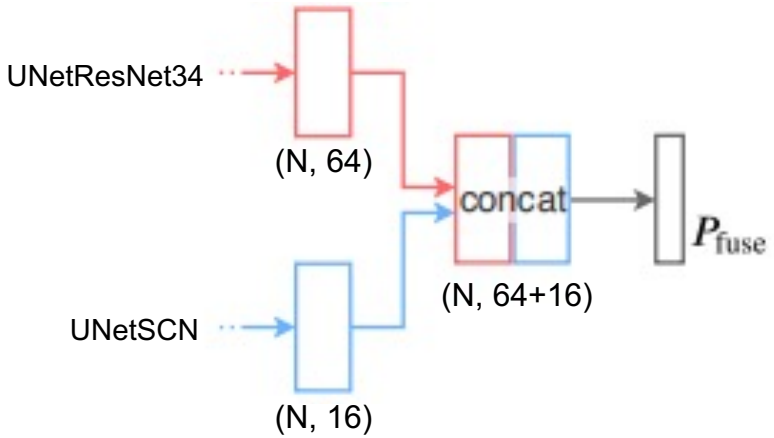


baseline_usa_v1:

- batch_size=8(the same as xmuda)
- $L = L_{seg}$



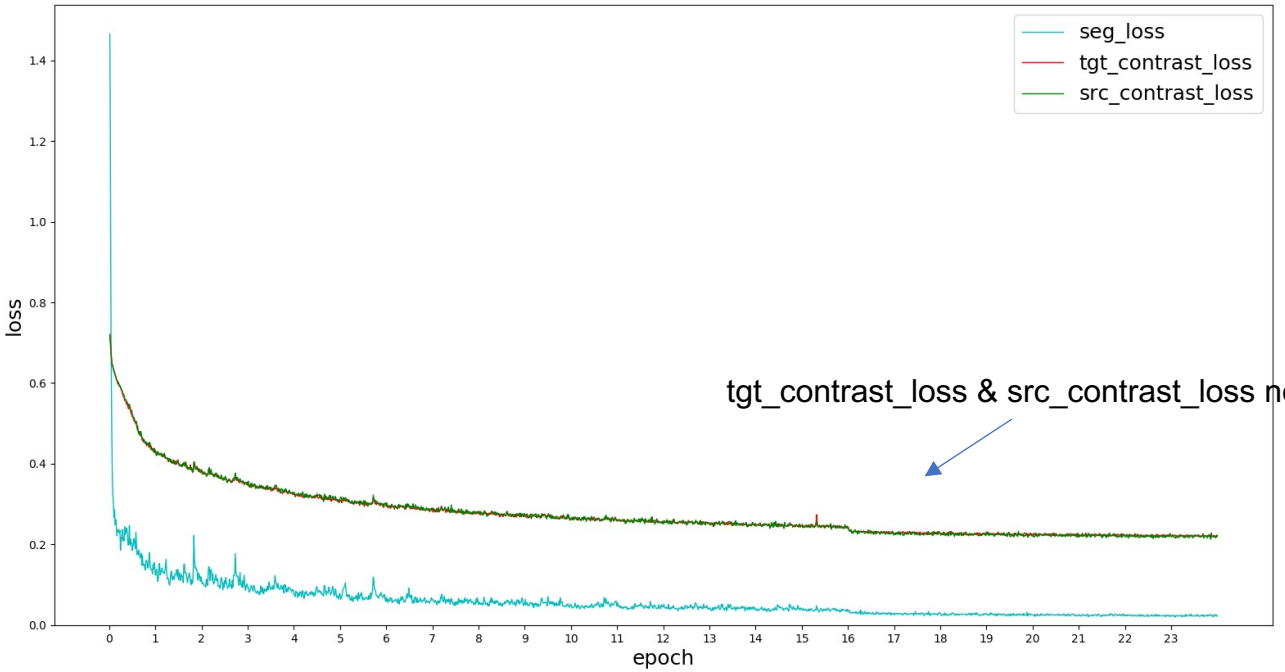




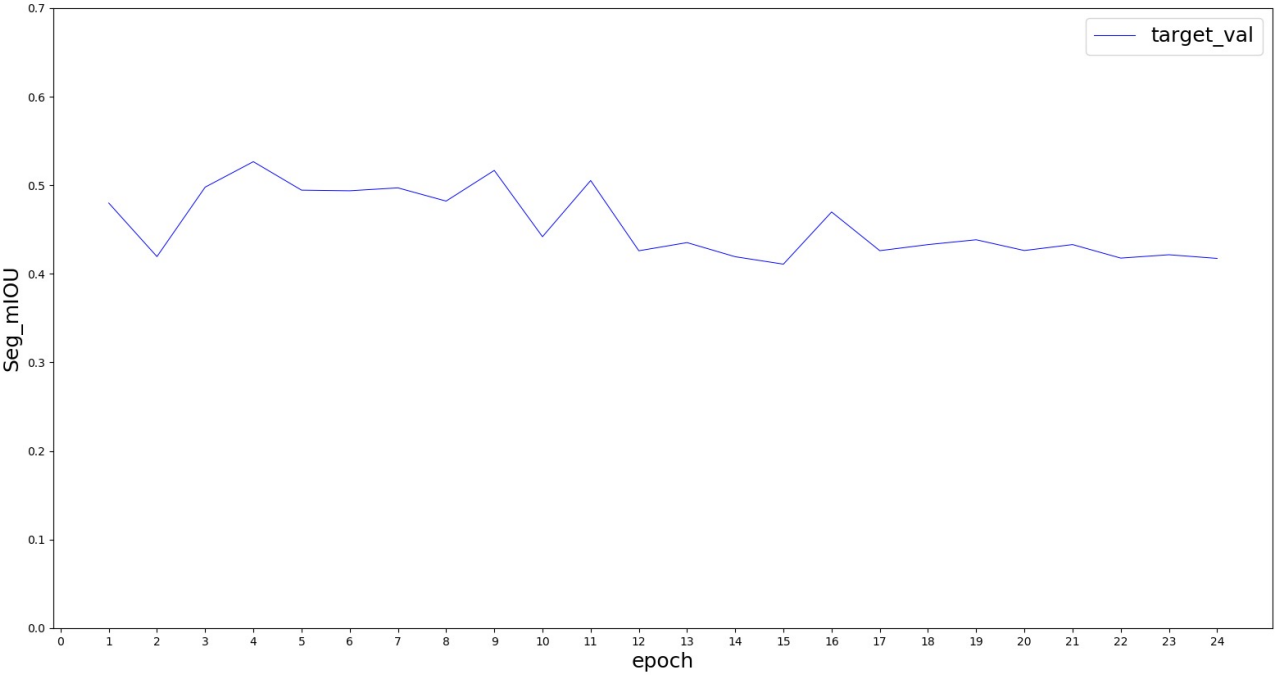
Vanilla Fusion(Baseline)

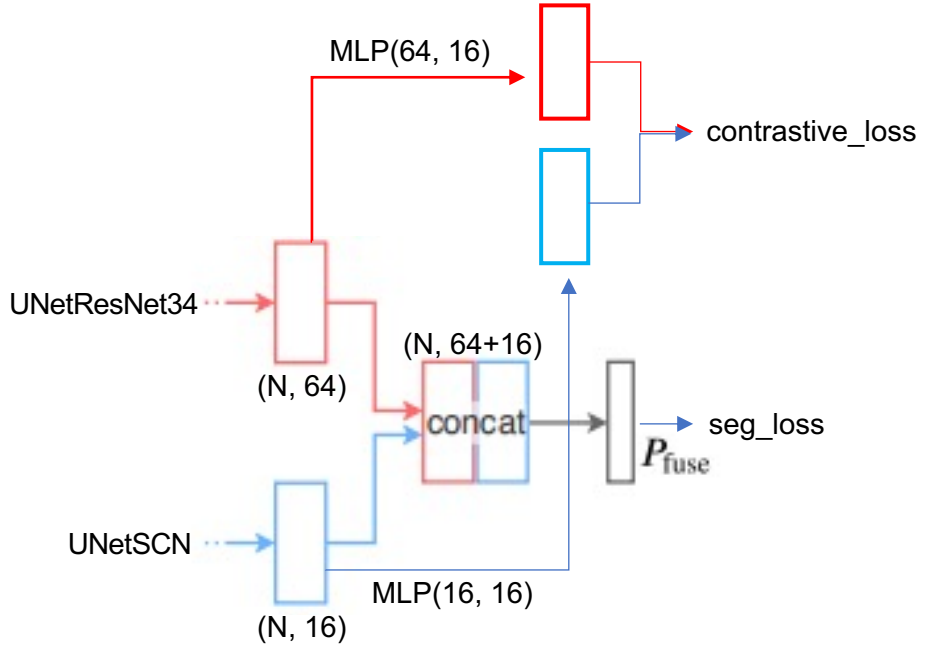
baseline_usa_v0:

- batch_size=4
- $L = L_{seg}$



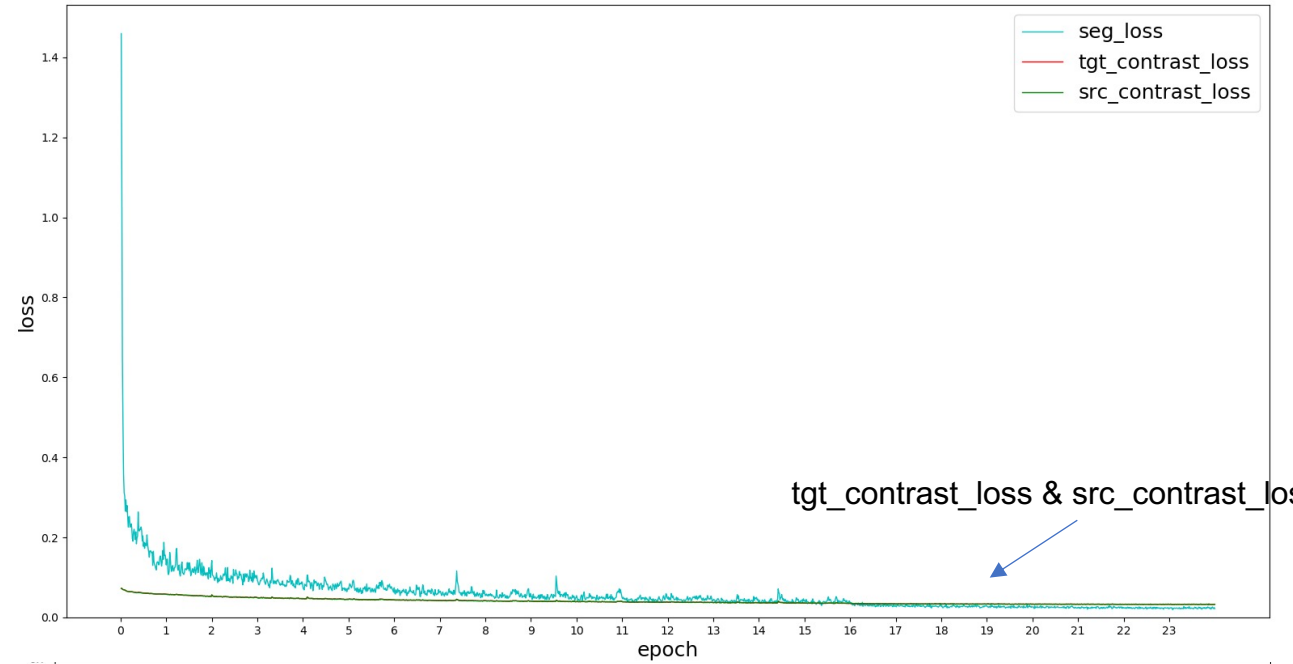
tgt_contrast_loss & src_contrast_loss nearly coincide



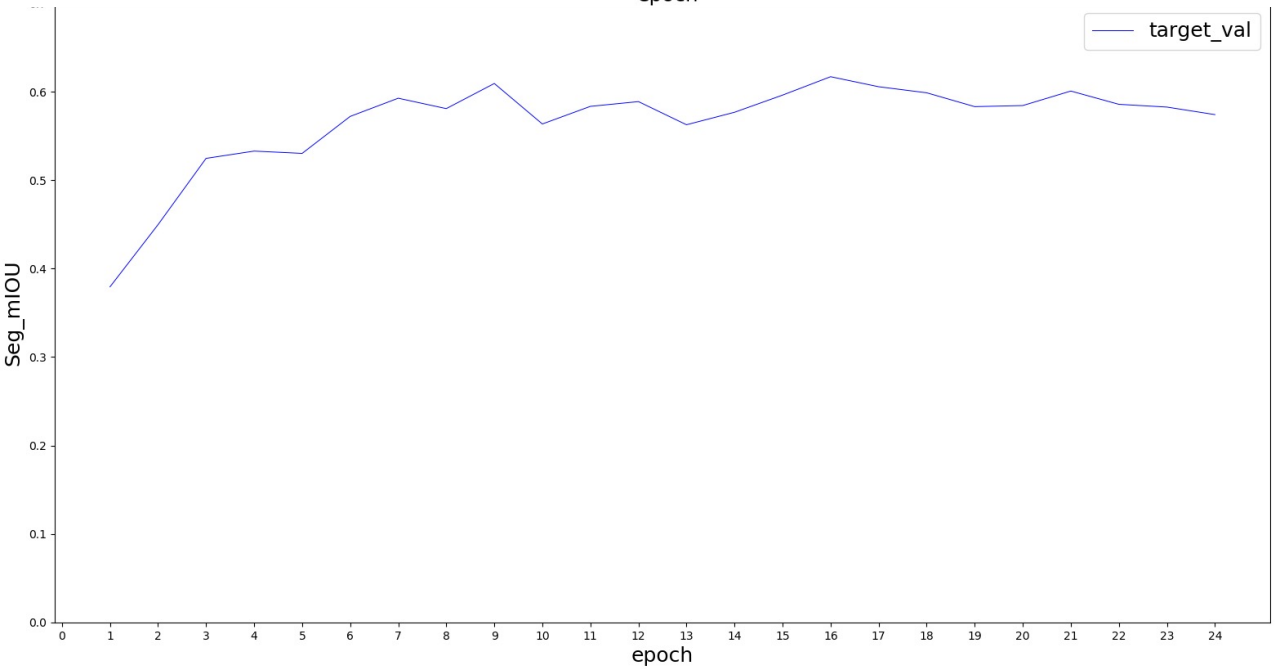


contrast_usa_v0:

- $\boldsymbol{\lambda = 0.1}$
- batch_size=8
- $L = L_{seg} + \lambda * (L_{src\_contrast} + L_{tgt\_contrast})$



tgt_contrast_loss & src_contrast_loss nearly coincide



contrast_usa_v1:

- $\boldsymbol{\lambda = 0.01}$
- batch_size=8
- $L = L_{seg} + \lambda * (L_{src\_contrast} + L_{tgt\_contrast})$

03/16/2021

- Xmuda new baseline (trained with new seg labels)
- “Multi-modal Multi-Task fusion Ver2” baseline
- Move xMuda network structure to our framework; train vanilla fusion

xMuda, reproduce results with new seg labels						
Train/test	USA/Singapore			Day/night		
Segmentation (mIoU)	Xmuda (dual head, KL_div)	Baseline (train on source)	Oracle (train on target)	Xmuda	Baseline	Oracle
2D	57.04	54.03	70.83	49.62	38.81	39.30
3D	53.57	48.41	65.55	45.55	43.75	46.84
2D+3D	62.53	62.10	75.40	52.99	48.63	43.39

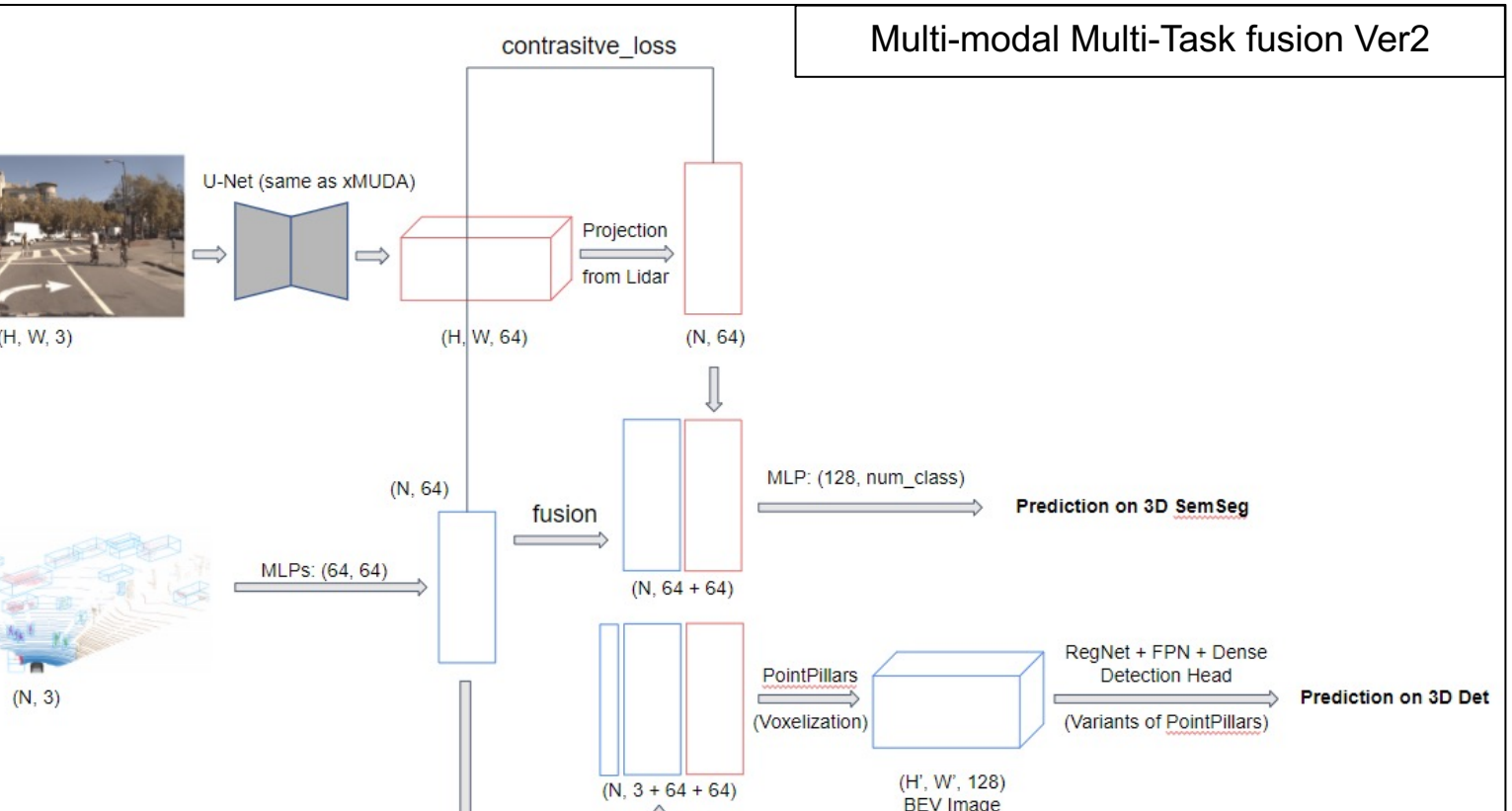
xMuda, old results showed in paper
(seg label obtained by marking points in bounding boxes)

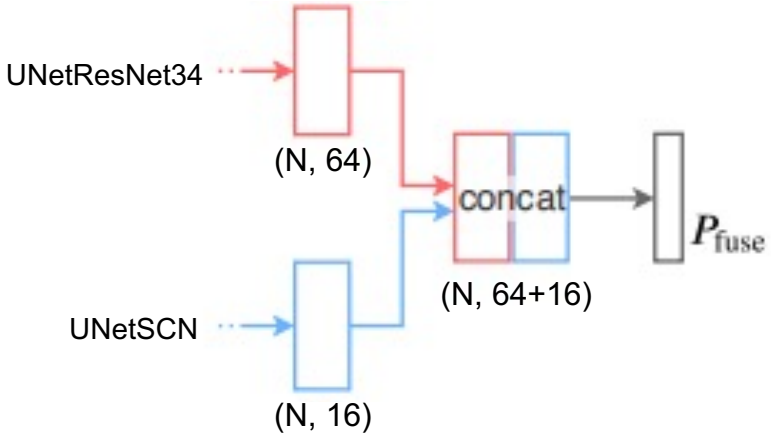
	USA/Singapore			Day/Night			A2D2/SemanticKITTI		
Method	2D	3D	softmax avg	2D	3D	softmax avg	2D	3D	softmax avg
Baseline (source only)	53.4	46.5	61.3	42.2	41.2	47.8	36.0	36.6	41.8
Deep logCORAL [21]	52.6	47.1	59.1	41.4	42.8	**51.8**	35.8*	39.3	40.3
MinEnt [29]	53.4	47.0	59.7	44.9	43.5	51.3	38.8	38.0	42.7
PL [17]	55.5	51.8	61.5	43.7	45.1	48.6	37.4	44.8	47.7
xMUDA	59.3	52.0	62.7	46.2	44.2	50.0	36.8	43.3	42.9
$xMUDA_{PL}$	**61.1**	**54.1**	**63.2**	**47.1**	**46.7**	50.8	**43.7**	**48.5**	**49.1**
Oracle	66.4	63.8	71.6	48.6	47.1	55.2	58.3	71.0	73.7

* Trained with batch size 6 instead of 8 to fit into GPU memory.

in/test: sa/Sng	Ver1 (separate encoders for points)	Multi-modal Multi-Task fusion Ver2 (shared encoder for points)		
	Baseline	Baseline (train on source)	Task_loss + lambda * Contrast_loss (tried a few hyperparameters)	
			Train from scratch	Resume from baseline model
nentation mIoU)	60.69	**62.48**	48.80	57.63
tection mAP)	51.29	45.02	37.15	41.59

Xmuda, new seg labels			
Train/test Usa/Sng	Xmuda (dual head, KL_div)	Baseline (separate two streams)	Vanilla fusion
2D	57.04	54.03	
3D	53.57	48.41	
2D+3D	**62.53**	**62.10**	**63.74**





Vanilla Fusion(Baseline)