APPENDIX A
COMPLEXITY ANALYSIS

The complexity of the main components of ASTCL is analyzed using the complexity theory, including ECG augmentations, encoder, discriminator, transformer, projector and predictor. In the pre-training stage, the time complexity of the core steps of these components are $\mathcal{O}(T\times C)$, $\mathcal{O}(ks\times K\times D^2)$, $\mathcal{O}(D^2)$, $\mathcal{O}(K\times M^2+K^2\times M)$ and $\mathcal{O}(H^2)$, and their space complexity are $\mathcal{S}(T\times C)$, $\mathcal{S}(ks\times D^2+K\times D)$, $\mathcal{S}(D^2+1)$, $\mathcal{S}(K^2+K)$ and $\mathcal{S}(H^2+H)$, where ks is the kernel size of the encoder, and the complexity of the projector and predictor can be regarded as equal. However, in the above components of ASTCL, only the encoder works in the fine-tuning stage and the testing stage. Thus, the time complexity is $\mathcal{O}(ks\times K\times D^2)$, and the space complexity is $\mathcal{S}(ks\times D^2+K\times D)$.

Because the encoder of ASTCL can be replaced, when the pre-trained encoder is used for clinical diagnosis, only the architecture of the encoder and the size of ECG signal will affect the diagnosis efficiency, while ASTCL will not increase time cost and memory cost in the diagnosis process.

APPENDIX B
DESCRIPTION OF DATASET

This section replenishes the description of selected datasets in the experiments. We use 4 ECG benchmark datasets to carry out the multi-class classification task and the multi-label classification task. To verify the performance of the model pre-trained by ASTCL in clinical diagnosis task, we establish a clinical ECG dataset in the experiment.

Chapman [58] was published in 2020, which contains 12-lead data with 500Hz from 10,646 patients of 10 seconds length. This dataset is multi-class dataset, including Sinus Bradycardia (SB), Sinus Rhythm (SR), Atrial Fibrillation (AFIB), Sinus Tachycardia (ST), Atrial Flutter (AF), Sinus Irregularity (SI), Supraventricular Tachycardia (SVT), Atrial Tachycardia (AT), Atrioventricular Node Reentrant Tachycardia (AVNRT), Atrioventricular Reentrant Tachycardia (AVRT) and Sinus Atrium to Atrial Wandering Rhythm (SAAWR). According to the suggestions of the literature [**?**], we combine ST, SVT, AT, AVNRT, AVRT and SAAWR into GSVT, classify AFIB as AF, and SI as SR. Hence, the categories in Chapman are AF, GSVT, SB and SR.

PTB-XL [59] was released in 2020, including 21,837 ECG signals of 18,885 patients. This dataset has 12 leads, whose sampling time is 10 seconds and the sampling frequency is 100Hz and 500Hz. PTB-XL is multi-label dataset with 5 categories, which includes Normal ECG (NORM), Myocardial Infarction (MI), Conduction Disturbance (CD), Hypertrophy (HYP), ST-T Segment Changes (STTC) and 71 subcategories. To enrich the multi-class classification task, we customize the PTB-XL to multi-class dataset. The data outside 5 categories or data with multiple labels are removed, and NORM records are reduced to 60 percent. In the experiment, the number of selected records in PTB-XL is 13,344.

CODE [60] was released in 2020 and consists of 2,322,513 12-lead ECG records from 1,676,384 different patients, whose sampling rate is 400Hz, and the recording duration of signals is 7 seconds to 10 seconds. This dataset is multi-label dataset, which have 7 categories, such as First-degree Atrioventricular Block (1dAVb), Right Bundle Branch Block (RBBB),

TABLE VIII
MAIN COMPONENTS DETAILS USED FOR ALL EXPERIMENTS

(a) Encoder

	Components	Dimension
Block 1	Conv1D	$ks=8$, $D_{in}=12$, $D_{out}=16$
	BatchNorm1d	
	ReLU	
	MaxPool1d	$ks=2$
	Dropout	$p=0.1$
Block 2	Conv1D	$ks=8$, $D_{in}=16$, $D_{out}=32$
	BatchNorm1d	
	ReLU	
	MaxPool1d	$k=2$
Block 3	Conv1D	$ks=8$, $D_{in}=32$, $D_{out}=64$
	BatchNorm1d	
	ReLU	
	MaxPool1d	$ks=2$
Block 4	Conv1D	$ks=8$, $D_{in}=64$, $D_{out}=128$
	BatchNorm1d	
	ReLU	
	MaxPool1d	$ks=2$

(b) Transformer

	Components	Dimension
Embedding	Linear	$M_{in}=128$, $M_{out}=100$
Block 1∼4	LayerNorm	
	Multi-head Attention	$head=4$, $M_{in}=100$
	LayerNorm	
	Linear	$M_{in}=100$, $M_{out}=64$
	ReLU	
	Dropout	$p=0.1$
	Linear	$M_{in}=64$, $M_{out}=100$
	Dropout	$p=0.1$

(c) Discriminator

	Components	Dimension
Connecting	Linear	$D_{in}=33\times128$, $D_{out}=128$
Block 1	Linear	$D_{in}=128$, $D_{out}=64$
	ReLU	
Block 2	Linear	$D_{in}=64$, $D_{out}=32$
	ReLU	
Block 3	Linear	$D_{in}=32$, $D_{out}=16$
	ReLU	
Block 4	Linear	$D_{in}=16$, $D_{out}=2$
	Sigmoid	

(d) Projector

Components	Dimension
Linear	$H_{in}=100$, $H_{out}=64$
BatchNorm1d	
ReLU	
Linear	$H_{in}=64$, $H_{out}=32$

(e) Predictor

Components	Dimension
Linear	$H_{in}=32$, $H_{out}=8$
BatchNorm1d	
ReLU	
Linear	$H_{in}=8$, $H_{out}=32$

TABLE IX
THE F1 SCORE OF NOISE EVALUATION IN ABLATION STUDY

Dataset	Chapman			PTB-XL		
SNR	2dB	5dB	10dB	2dB	5dB	10dB
ASTCL (-AG)	86.87±0.81	88.08±1.36	88.54±0.94	55.41±0.73	55.93±0.36	56.39±0.41
ASTCL (-RN)	**87.51±0.57**	**88.85±1.13**	**89.09±1.21**	**56.16±1.41**	**56.85+0.67**	**57.01±0.37**
ASTCL	**87.77±0.75**	**88.78±0.71**	**89.25±1.02**	**56.58±1.04**	**56.99±0.56**	**57.36±0.92**
Dataset	CODE			CPSC2018		
SNR	2dB	5dB	10dB	2dB	5dB	10dB
ASTCL (-AG)	80.87±0.97	82.92±0.54	83.20±0.65	58.19±1.04	60.36±1.24	60.85±1.33
ASTCL (-RN)	**81.15±0.59**	**83.41±0.34**	**83.94±0.28**	**61.07±0.61**	**62.53±1.48**	**63.38±0.74**
ASTCL	**81.67±0.70**	**83.87±0.51**	**84.13±0.59**	**61.56±1.17**	**62.62±0.83**	**63.34±1.26**

Left Bundle Branch Block (LBBB), Sinus Bradycardia (SB), Sinus Tachycardia (ST), Atrial Fibrillation (AF) and No abnormalities (Norm). To maintain a similar standards with other datasets, we select the 10 seconds from exams_part0 to exams_part3 of CODE as experimental data, with 12,559 records in total.

CPCS2018 [61] was published in 2018, which includes 6,877 patients with 12-lead records. The sampling rate of the data is 500Hz and the recording length is 6 seconds to 60 seconds. This dataset included 9 kinds of arrhythmias, such as Atrial fibrillation (AF), First-degree Atrioventricular Block (I-AVB), Left Bundle Branch Block (LBBB), Normal Sinus Rhythm (NSR), Premature Atrial Contraction (PAC), Premature Ventricular Contraction (PVC), Right Bundle Branch Block (RBBB), ST-segment Depression (STD) and ST-segment Elevated (STE). To standardize the length of the records, the length of signal longer than 10 seconds is cut according to an integer multiple of 10 seconds, and delete those less than 10 seconds.

Clinical myocardial infarction (CMI) dataset is a clinical ECG dataset, which collected under the Cooperative Innovation Center for Internet Healthcare of Zhengzhou University, and labeled by professional cardiologists. CMI consists of 10,336 12-lead records from 7,317 patients. The recording duration of each data is 10 seconds, and their sampling frequency is 500Hz. This dataset is multi-label dataset, which includes 4 main Myocardial Infarction categories, such as Anterior Myocardial Infarction (AMI), Inferior Myocardial Infarction (IMI), Lateral Myocardial Infarction (LMI), Posterior Myocardial Infarction (PMI). Among them, AMI contains 4,728 records, and IMI has 2,823 records. The number of LMI and PMI are 320 and 304 respectively. In addition, there are 3,167 Normal (Norm) data in CMI.

APPENDIX C
IMPLEMENTATION DETAILS

This section supplements the implementation details of data setup and hyperparameter setup, and introduces the specifics of experimental environment and platform in development.

A. Data Setup

To reduce the impact of instrument and personal differences, the data amplitude of five datasets is normalized between 0 and 1 by Z-score normalization [47], and all data is resample to 250Hz (limited by mechanism, CLOCS uses 500Hz data). As for the selection of data length, we take the data with length T = 2,500 as the experimental data. If the data length is not equal to 2,500, we will discard data with $T <$ 2,500 or segment data with $T >$ 2,500.

Each training set, validation set, and testing set are randomly divided from each dataset according to the proportion of 60%, 20% and 20%. We repeat these experiments five times with 5 different seeds respectively and analyze the mean and standard deviation of each experiment. Whether in pre-training stage, fine-tuning stage and testing stage, the batchsize and epochs are all set to 128 and 100.

B. Hyperparameter Setup

The details of the encoder, transformer, discriminator, projector and predictor of we used are defined as shown in Table VIII. The output dimension of encoder, transformer and projector is set D = 128, M = 100 and H = 32 respectively. We employ the Adam optimizer to update the model parameters, and set the learning rate to η = 3e-4 and weight decay to $\beta 1$ = 0.9 and $\beta 2$ = 0.99. In ECG augmentations, the SNR used in this paper is defined as μ = 5dB. In spatiotemporal prediction task, the past segment length τ is set to $0.6K$, which is the same as in the study of [13]. To keep ASTCL achieve the best performance, the weight w_1, w_2 and w_3 of loss function $\mathcal{L}_F$ are all set to 1 in the experiment.

In settings of baseline methods, the temperature parameter of NT-Xent loss function used in SimCLR, CLOCS and TS-TCC is all set to 0.2. Furthermore, the decay rate in BYOL is define as 0.90. To ensure fairness of experiment, each baseline method uses the same input, data augmentation, encoder, projector, optimizer and classifier as ASTCL.

C. Environment & Platform Details

In the aspect of the implementation environment, the development of ASTCL and the replication of baseline methods are bulit using Python 3.6.2 on the operating system Ubantu 16.4. The selected machine learning framework is PyTorch 1.4, and the used CUDA version is 10.1. The experimental platform is equipped with Inter Xeon Silver 4114@2.20GHz CPU, and its GPU is NVIDIA GeForce RTX 2080Ti. The memory of the platform is 128GB DDR4 DRAM.

TABLE X
THE F1 SCORE OF CATEGORY EVALUATION IN ABLATION STUDY

(a) Chapman

Category	ASTCL (-AG)	ASTCL (-RN)	ASTCL
AFIB	82.87±3.43	81.13±1.16	**82.99±2.45**
GSVT	**87.58±1.87**	86.31±0.34	87.43±1.01
SB	96.93±0.26	97.04±0.82	**96.96±0.33**
SR	90.68±1.17	89.96±2.28	**90.80±0.94**
Total	89.46±0.88	88.61±0.64	**89.54±1.11**

(b) PTB-XL

Category	ASTCL (-AG)	ASTCL (-RN)	ASTCL
CD	72.54±0.56	71.39±0.93	**72.61±0.61**
HYP	6.02±1.42	5.74±1.98	**6.20±1.20**
MI	61.01±0.69	59.25±2.32	**61.39±1.48**
NORM	82.85±0.41	**83.17±0.72**	83.06±0.54
STTC	**64.73±1.76**	63.28±1.02	63.76±1.05
Total	**57.43±0.91**	56.57±0.65	57.40±0.88

(c) CODE

Category	ASTCL (-AG)	ASTCL (-RN)	ASTCL
1dAVb	67.37±2.31	64.85±3.76	**68.33±2.34**
RBBB	93.53±0.88	92.94±0.58	**93.75±0.31**
LBBB	**90.64±0.61**	89.92±1.07	90.38±1.32
SB	85.13±0.83	83.63±0.90	**85.41±1.33**
ST	**90.86±1.29**	90.80±1.06	90.76±0.72
AF	70.82±1.05	69.03±1.38	**72.27±1.55**
Norm	94.12±0.51	93.84±0.36	**94.32±0.15**
Total	84.64±0.57	83.57±0.68	**85.03±0.42**

(d) CPSC2018

Category	ASTCL (-AG)	ASTCL (-RN)	ASTCL
PVC	59.27±0.45	58.41±1.35	**59.62±0.96**
AF	76.48±1.03	75.53±1.57	**76.80±1.31**
LBBB	90.59±0.53	**90.75±0.79**	90.69±0.44
STE	42.01±4.64	39.58±3.43	**43.24±3.62**
IAVB	69.75±3.07	65.53±5.29	**71.71±2.98**
PAC	**22.91±2.86**	21.38±2.13	22.29±2.11
NSR	**61.25±2.02**	59.87±2.46	60.96±1.19
STD	58.73±2.25	58.36±2.83	**58.76±2.31**
RBBB	89.02±0.87	88.97±0.15	**89.11±0.64**
Total	63.41±0.73	62.04±1.15	**63.69±0.69**

(e) CMI

Category	ASTCL (-AG)	ASTCL (-RN)	ASTCL
AMI	95.66±0.54	95.30±0.86	**95.75±0.31**
IMI	**88.28±0.80**	87.92±0.74	88.19±1.00
LMI	38.33±3.95	35.64±5.49	**40.61±4.03**
PMI	31.73±3.46	26.25±5.18	**32.30±4.01**
Norm	96.45±0.45	95.99±0.61	**96.52±0.33**
Total	70.09±1.03	68.22±1.62	**70.61±1.90**

APPENDIX D
SUPPLEMENT OF ABLATION STUDY

In ASTCL, we employ adversarial game task and only using patient-level positive pairs to improve anti-perturbation ability and better learn category representations. To further verify the contribution of these solutions to ASTCL, we extend noise evaluation and category evaluation in the ablation study. As described in section V.G Ablation Study, ASTCL (-AG) means ASTCL without adversarial game task, and ASTCL (-RN) means ASTCL without removing negative pairs operation. Next, we will introduce experiments in detail, and the best results are marked in **black** and the second-best in **red**.

A. Noise Evaluation in Ablation Study

It is mentioned in the introduction that we proposed a self-supervised task (i.e., adversarial game) for discarding noise representations. To verify the effect of adversarial game task, we use ASTCL (-AG), ASTCL (-RN) and ASTCL to conduct noise evaluation experiment. Same as the experimental setup in section V.C Noise Evaluation, all frameworks are pre-trained on 4 ECG banchmark datasets. Then, we add baseline drift noise, muscle artifacts noise and power frequency noise to the original signal using the SNR of 2dB, 5dB and 10dB based on general noise stress evaluation [63]. Lastly, the pre-trained model is fine-tuned by 50% labeled noised ECG data to perform classification of noised data.

Table IX shows the F1 scores of experimental frameworks in noise evaluation experiment. Among the three frameworks, the complete ASTCL outperforms other frameworks, it achieves the top1 in 10 of 12 groups of experiment. But ASTCL (-AG) performs the worst. Especially on the CPSC2018 dataset, compared with the complete ASTCL, the F1 score of ASTCL (-AG) is reduced by 3.37%, 2.26% and 2.49%. This means that the robustness of ASTCL to noise becomes weak after the lack of adversarial game task. Besides, the performance of ASTCL (-RN) is also lower than the complete ASTCL, but it is still significantly better than ASTCL (-AG). This confirms the view that only using patient-level positive pairs can also bring gains to the performance of the model, but its contribution to improving the robustness of ASTCL to noise is not critical. The adversarial game task is the key to improve the anti-perturbation ability of the model.

B. Category Evaluation in Ablation Study

To learn the category representations better, we only employ patient-level positive pairs, and alternately utilize the predictor and stop gradient of projection to replace negative pairs. In proving the validity of only using patient-level positive pairs, the category evaluation experiment is extended in ablation study to test ASTCL (-AG), ASTCL (-RN) and ASTCL. These contrastive learning frameworks are pre-trained on five datasets respectively, and fine-tuned by 50% labeled data to identify the categories of data, which is the same as the experimental process in section V.D Category Evaluation.

In Table X, we show the F1 scores of ASTCL (-AG), ASTCL (-RN) and ASTCL in category evaluation experiment. We observe that the complete ASTCL's ability to identify categories is the most stable on each dataset, while the performance of ASTCL (-RN) is lower than the other two frameworks. Referring to the complete ASTCL, the total F1 score of ASTCL (-RN) in the five datasets decreases by 0.93%, 0.83%, 1.46%, 1.65% and 2.39% respectively. Particularly, in the 1dAVb, STE, IAVB, LMI and PMI, the performance of ASTCL (-RN) is significantly reduced. Among them, ASTCL (-RN) decreases by 6.05% in PMI. These experimental results show that removing negative pairs operation can indeed improve the accuracy of category recognition. Therefore, this

TABLE XI
THE F1 SCORE OF NOISE EVALUATION IN AVAILABILITY STUDY

Dataset	Chapman			PTB-XL		
SNR	2dB	5dB	10dB	2dB	5dB	10dB
TS-TCC [13]	86.85±1.05	87.76±0.96	88.27±0.94	55.29±1.47	56.23±1.51	56.27±1.91
TS-TCC (+AG)	**87.54±0.29**	**88.52±0.59**	**88.93±0.62**	**56.14±0.65**	**56.91±1.16**	**57.03±0.82**
CLOCS [16]	86.49±1.13	87.41±0.82	87.63±0.99	53.56±0.83	55.31±0.88	55.44±0.61
CLOCS (+AG)	**87.38±1.11**	**88.16±0.63**	**88.51±0.34**	**55.74±0.49**	**56.85±0.68**	**56.79±0.73**
Dataset	CODE			CPSC2018		
SNR	2dB	5dB	10dB	2dB	5dB	10dB
TS-TCC [13]	79.77±0.66	81.47±0.75	81.89±0.60	58.63±1.19	59.81±1.02	60.48±0.91
TS-TCC (+AG)	**80.82±0.33**	**82.64±0.50**	**83.17±0.35**	**60.31±0.83**	**61.73±1.23**	**62.22±1.32**
CLOCS [16]	79.46±0.38	80.29±0.83	80.65±1.05	54.76±1.35	57.15±0.56	58.39±0.86
CLOCS (+AG)	**80.18±0.32**	**80.68±0.28**	**81.19±0.29**	**56.14±0.86**	**58.72±1.19**	**59.57±0.98**

TABLE XII
THE F1 SCORE OF CATEGORY EVALUATION IN AVAILABILITY STUDY (BENCHMARK DATASETS)

(a) Chapman

Category	TS-TCC [13]	TS-TCC (+RN)	CLOCS [16]	CLOCS (+RN)
AFIB	81.97±2.05	**82.62±3.17**	80.22±1.74	**81.07±1.16**
GSVT	86.12±1.47	**87.34±1.86**	86.00±0.43	**86.82±0.91**
SB	**96.91±0.22**	96.83±0.35	96.52±0.28	**96.78±0.18**
SR	90.29±0.75	**90.71±1.54**	88.82±1.05	**89.14±1.78**
Total	88.82±0.95	**89.38±1.59**	87.89±0.84	**88.45±0.29**

(b) PTB-XL

Category	TS-TCC [13]	TS-TCC (+RN)	CLOCS [16]	CLOCS (+RN)
CD	71.58±0.84	**72.42±0.51**	69.37±1.22	**70.48±0.87**
HYP	6.50±1.34	**7.57±0.92**	4.40±1.67	**5.59±1.06**
MI	58.04±1.39	**59.39±1.27**	**57.33±2.49**	57.24±2.78
NORM	**82.26±0.66**	82.14±0.35	**82.79±0.29**	82.75±1.36
STTC	63.61±1.25	**64.09±0.61**	64.95±0.61	**65.37±0.94**
Total	56.42±0.87	**57.12±0.49**	55.77±0.88	**56.29±0.58**

(c) CODE

Category	TS-TCC [13]	TS-TCC (+RN)	CLOCS [16]	CLOCS (+RN)
1dAVb	61.29±1.67	**63.79±2.01**	57.44±4.05	**60.94±3.63**
RBBB	92.92±0.69	**93.69±0.81**	92.52±0.39	**93.31±0.52**
LBBB	**89.35±0.68**	89.27±0.37	89.24±0.59	**89.44±0.43**
SB	84.19±1.77	**86.45±1.36**	82.55±0.53	**83.02±1.14**
ST	89.77±0.41	**90.18±0.62**	**88.27±0.72**	88.25±0.98
AF	63.19±2.11	**66.54±1.31**	57.92±2.69	**61.97±1.56**
Norm	93.76±0.40	**94.29±0.37**	93.03±0.49	**93.79±0.62**
Total	82.07±0.55	**83.60±0.59**	80.14±1.00	**81.82±0.87**

(d) CPSC2018

Category	TS-TCC [13]	TS-TCC (+RN)	CLOCS [16]	CLOCS (+RN)
PVC	58.73±1.81	**59.58±1.56**	57.48±1.15	**58.39±2.54**
AF	75.81±2.09	**77.35±1.74**	70.56±1.42	**73.82±1.09**
LBBB	**90.32±1.01**	90.16±1.88	90.19±1.99	**90.43±2.26**
STE	40.34±2.34	**42.28±3.93**	38.13±5.57	**41.97±4.31**
IAVB	64.19±2.73	**66.38±3.04**	60.94±3.42	**63.22±1.23**
PAC	22.79±2.41	**23.93±1.51**	20.15±2.13	**23.88±2.68**
NSR	60.73±3.94	**61.75±1.87**	57.07±3.69	**59.16±2.32**
STD	**58.92±1.35**	58.84±1.54	56.04±3.30	**56.60±2.75**
RBBB	88.76±0.91	**88.81±0.68**	**87.51±0.53**	87.42±0.33
Total	62.29±1.07	**63.23±0.64**	59.79±1.12	**61.65±1.30**

TABLE XIII
THE F1 SCORE OF CATEGORY EVALUATION IN AVAILABILITY STUDY (CMI DATASET)

Category	TS-TCC [13]	TS-TCC (+RN)	CLOCS [16]	CLOCS (+RN)
AMI	95.33±0.88	**96.58±0.83**	94.64±0.51	**95.65±0.36**
IMI	**88.23±0.59**	88.16±0.39	88.22±0.69	**88.52±0.53**
LMI	41.01±4.72	**42.72±3.55**	26.82±5.02	**32.38±4.29**
PMI	31.23±3.36	**33.01±2.82**	23.58±4.62	**29.87±4.34**
Norm	**96.29±0.16**	96.25±0.38	96.37±0.24	**96.44±0.41**
Total	70.26±1.28	**71.34±1.17**	66.62±1.78	**68.57±0.76**

proves our view that only using patient-level positive pairs is conducive to better learning category representations.

APPENDIX E
AVAILABILITY STUDY

Adversarial game task and only using patient-level positive pairs are improvements to the framework. To futher test whether these two solutions are useful in other frameworks, we conduct the availability study in this section, including noise evaluation and category evaluation. The best results are marked in **black**.

A. Noise Evaluation in Availability Study

To validate the availability and effectiveness of the adversarial game task, in noise evaluation experiment, we equip the adversarial module of ASTCL to TS-TCC [13] and CLOCS [16], which are the state-of-the-art contrastive learning frameworks in PTS field. We call the reformed TS-TCC and CLOCS as TS-TCC (+AG) and CLOCS (+AG). For fairness, the proposed ECG augmentations are adopted in the experiment for data augmentations of each contrastive learning framework. These experimental frameworks are first pre-trained on 4 ECG banchmark datasets, and fine-tuned by 50% labeled noised ECG data. The noised ECG data used in this experiment are the same as those used in the ablation study.

The results of this experiment are shown in the Table XI. We find that the F1 scores of TS-TCC (+AG) and CLOCS (+AG) are always higher than that of TS-TCC and CLOCS, regardless of any dataset or any SNR. Especially on CPSC2018 dataset, the improvement brought by the adversarial game task is most obvious. TS-TCC (+AG) increases by 1.68%, 1.92% and 1.74% compared with the original TS-TCC framework, and CLOCS (+AG) also increases by 1.38%, 1.57% and 1.18% compared with CLOCS, which means that adversarial game task is also applicable to TS-TCC and CLOCS, and effectively improve their anti-perturbation ability.

B. Category Evaluation in Availability Study

Using patient-level positive pairs and removing negative pairs are the key for ASTCL to improve the ability of learning category representations. To test the performance of this solution in other frameworks, we integrate this operation into TS-TCC [13] and CLOCS [16] of the same field as ours to perform category evaluation. Specifically, The predictor and stop-gradient operation are added into TS-TCC and CLOCS, and $\mathcal{L}_C$ of ASTCL is used as the loss function to optimize TS-TCC and CLOCS. These modified frameworks are defined TS-TCC (+RN) and CLOCS (+RN). In the experiment, we still use ECG augmentations to transform the data used by the experimental frameworks, and use five datasets to carry out category evaluation.

The pre-trained model is fine-tuned by 50% labeled data to identify categories. The F1 scores of the experimental frameworks in each category are shown in Table XII and Table XIII. We observe that this solution is still effective for TS-TCC and CLOCS. The F1 scores of TS-TCC (+RN) and CLOCS (+RN) in almost all categories are higher than the original TS-TCC and CLOCS. For example, the F1 scores of AFIB, HYP, 1dAVb, PAC and PMI is the worst of the original frameworks on five datasets, however, TS-TCC (+RN) and CLOCS (+RN) significantly improve performance. In 1dAVb, TS-TCC (+RN) increases F1 score from 61.29% to 63.79%. It is worth noting that in PMI, CLOCS (+RN) is 6.29% higher than CLOCS. This validates the viewpoint that only using patient-level positive pairs is still valid for other contrastive learning frameworks.