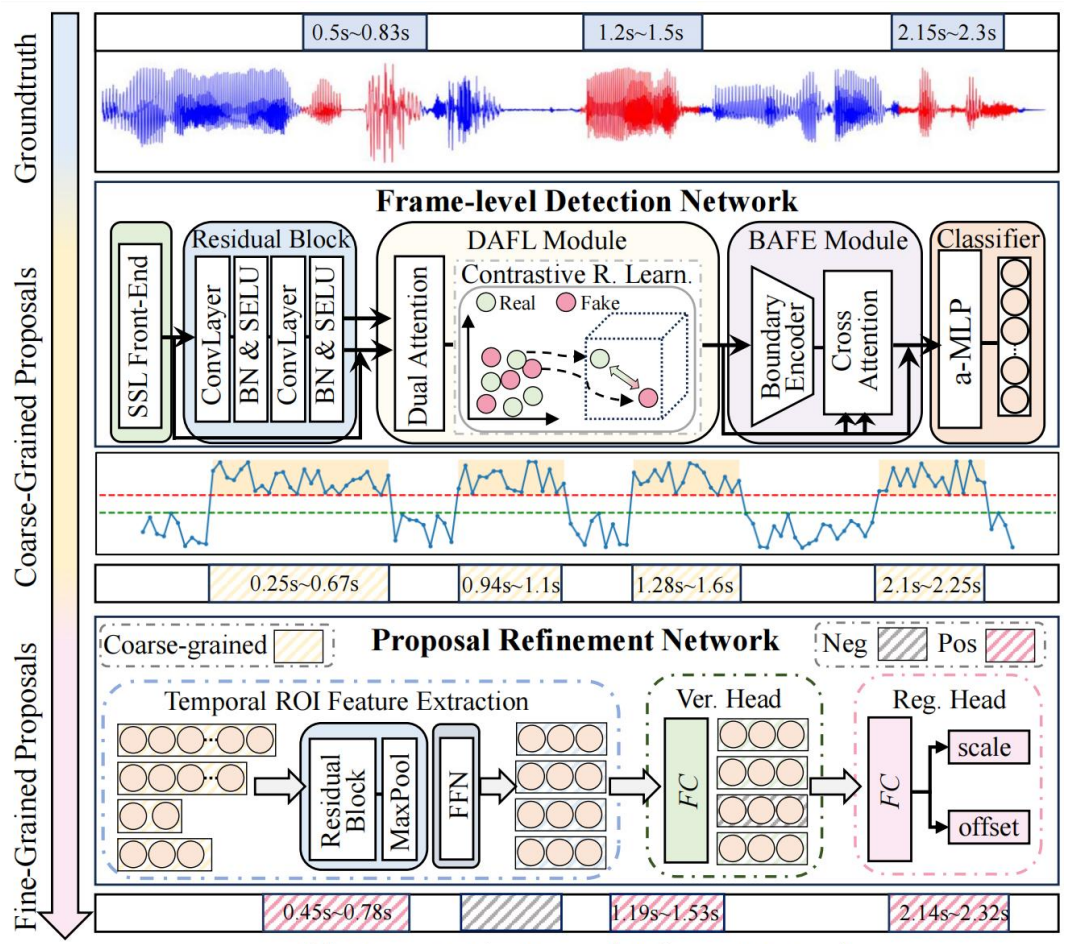
Coarse-to-Fine Proposal Refinement Framework for Audio Temporal Forgery Detection and Localization

24年 ACM MM（开源）

1. 目标：音频，伪造检测及时间定位
2. 数据集：LAV-DF（v+a），HAD（a），ASVS2019PS（a）
3. 方法：从粗到细的提案细化框架（CFPRF）：帧级检测网络（FDN，粗粒度，单独训练）+提案细化网络（PRN，细粒度，即插即用）
4. FDN网络：
   ①使用 SSL 模型（即 Wav2Vec2 - XLSR300M）提取前端特征；
   ②使用六个基于 CNN 的残差块来学习更高级的特征图；
   ③经过两个模块，经过 Classifier 得到帧特征和帧分数。
5. FDN包括两个模块：
   ①差异感知特征学习（DAFL）模块，扩大微小篡改引起的差异
   ②边界感知特征增强（BAFE）模块，捕获过渡边界的上下文信息，引导边界信息与时间特征之间的交互
6. PRN网络：输入是FDN中得到的帧特征和帧分数
   ①粗粒度提案
   ②经过池化编码器（两个基于 CNN 的残差块+一个最大池化层）
   ③采用两个基于 MLP的头部，预测的起始偏移量和持续时间偏移量
   ④得到细粒度提案



(a) Coarse-to-Fine Proposal Refinement Network

Coarse-to-Fine Proposal Refinement Framework for Audio Temporal Forgery Detection and Localization

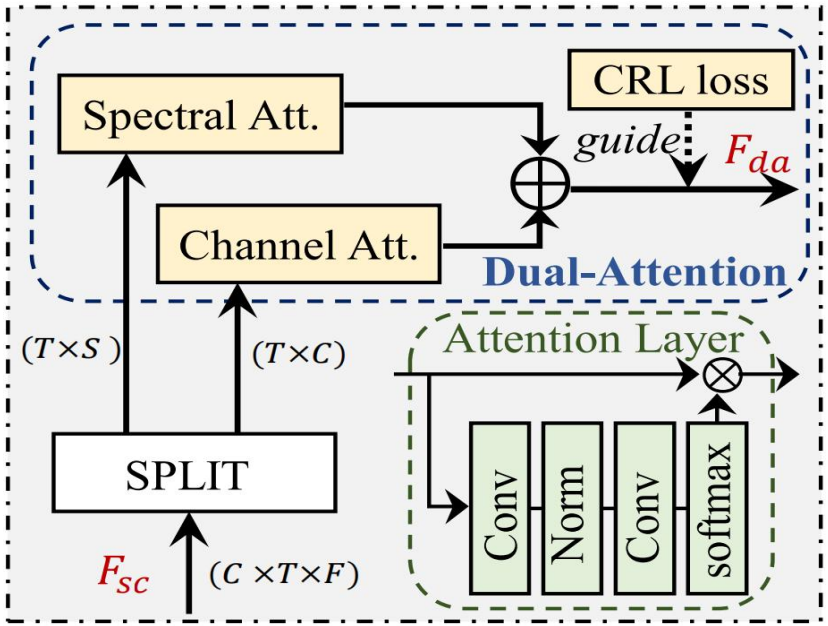
24年 ACM MM（开源）

7. FDN的两个模块：差异感知特征学习（DAFL）模块和边界感知特征增强（BAFE）模块

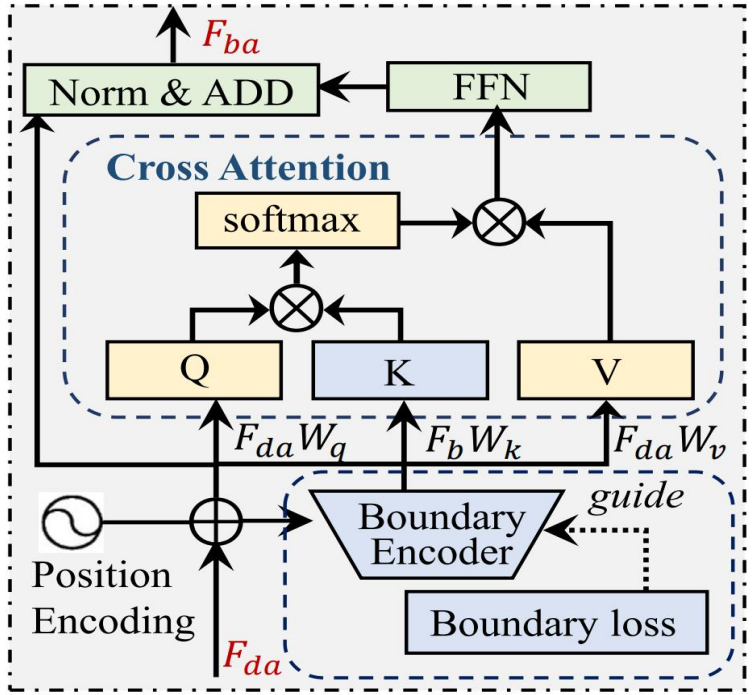
①DAFL：分离频谱S和通道特征C，经过AttentionLayer，双特征整合后对比学习，真真/假假（正对），真假（负对）

②BAFE：BoundaryEncoder+交叉注意力+FNN+Norm

BoundaryEncoder：aMLP编码器+全连接层进行下采样，得到边界概率。考虑标签不平衡，使用无超参损失函数：P2sGrad 的均方误差（MSE）。Classifier：使用 aMLP 解码器，得到帧级伪造概率分数，使用基于 P2sGrad 的 MSE 作为损失函数。



(b) Difference-Aware Feature Learning



(c) Boundary-Aware Feature Enhancement

Dataset	Method	EER ↓	AUC ↑	Pre↑	Rec↑	F1↑
HAD	IFBDN	0.35	**99.98**	99.92	99.65	99.78
	PSDL	0.18	99.97	99.96	99.82	99.89
	CFPRF	**0.08**	99.96	**99.98**	**99.92**	**99.95**
PS	IFBDN	9.68	95.70	93.72	90.32	91.99
	PSDL	12.47	93.30	91.82	87.53	89.62
	CFPRF	**7.41**	**96.97**	**95.23**	**92.59**	**93.89**
LAV-DF	IFBDN	1.07	99.88	99.94	98.93	98.93
	PSDL	0.82	**99.92**	99.95	99.18	**99.57**
	CFPRF	**0.82**	99.89	**99.95**	**99.18**	99.56

部分伪造检测结果。即对比的FDN网络

Coarse-to-Fine Proposal Refinement Framework for Audio Temporal Forgery Detection and Localization

24年 ACM MM（开源）

Dataset	Method	AP@0.5	AP@0.75	AP@0.9	AP@0.95	mAP	AR@1	AR@2	AR@5	AR@10	AR@20
HAD	BA-TFD	79.86	37.98	5.55	0.57	40.93	45.12	47.53	49.99	52.09	55.15
	BA-TFD+	88.26	70.69	37.83	7.39	64.83	67.49	68.44	69.06	69.39	70.15
	UMMAF	**99.98**	**99.86**	98.01	88.17	98.49	98.68	98.73	98.84	98.85	98.86
	IFBDN+PRN†	93.85	91.55	87.75	79.08	90.40	96.07	97.39	97.54	97.54	97.54
	IFBDN+PRN	99.30	98.02	95.30	89.47	97.12	97.08	97.53	97.53	97.53	97.53
	PSDL+PRN†	88.53	85.27	80.80	73.25	84.25	93.40	96.30	96.89	96.94	96.94
	PSDL+PRN	98.94	97.10	93.06	86.13	95.88	96.48	96.61	96.61	96.61	96.61
	CFPRF	99.77	99.60	**99.21**	**96.03**	**99.23**	**99.31**	**99.38**	**99.38**	**99.38**	**99.38**
PS	BA-TFD	13.65	4.91	1.06	0.63	6.15	8.04	11.03	15.41	19.14	23.64
	BA-TFD+	15.72	6.37	2.05	1.95	7.69	7.93	12.62	18.28	22.17	26.71
	UMMAF	52.99	31.89	17.69	9.04	33.09	17.37	28.49	39.57	47.55	55.53
	IFBDN+PRN†	43.84	34.79	27.10	22.53	34.92	18.72	33.30	53.87	60.99	62.22
	IFBDN+PRN	58.65	49.30	41.39	35.33	48.79	18.52	34.77	55.41	64.47	62.23
	PSDL+PRN†	46.63	38.19	31.13	26.94	38.42	**20.22**	35.16	56.86	64.97	66.52
	PSDL+PRN	54.25	46.47	40.57	36.70	46.68	19.70	**36.10**	**58.09**	65.40	66.51
	CFPRF	**66.34**	**55.47**	**48.05**	**40.96**	**55.22**	18.48	35.57	58.06	**65.47**	**66.53**
LAV-DF	BA-TFD	53.53	10.98	0.36	0.02	20.77	29.56	32.22	34.73	38.03	44.66
	BA-TFD+	83.78	51.99	6.13	0.46	49.32	52.78	54.97	57.21	58.41	60.04
	UMMAF	**97.29**	**95.67**	89.92	61.97	92.04	85.67	91.77	**94.89**	**95.64**	**96.14**
	IFBDN+PRN†	86.83	84.02	77.85	70.09	82.55	86.28	91.78	92.13	92.13	92.13
	IFBDN+PRN	94.02	92.49	89.42	84.60	91.69	86.62	92.12	92.16	92.16	92.16
	PSDL+PRN†	76.10	71.71	65.16	57.13	70.43	84.71	89.14	89.98	90.03	90.03
	PSDL+PRN	92.76	90.01	85.95	78.85	89.02	85.66	90.03	90.06	90.06	90.06
	CFPRF	94.52	93.47	**91.65**	**88.64**	**93.01**	**87.59**	**93.49**	93.51	93.51	93.51

8. PRN是即插即用的，配合现有的FDN网络进行对比；

9. 其中带 † 的是粗粒度提案