ICML
International Conference
On Machine Learning

Listenable Maps for Audio Classifiers

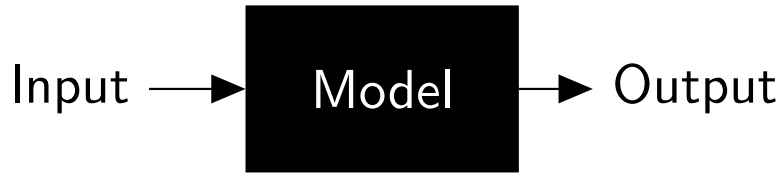
Francesco Paissan, Mirco Ravanelli, Cem Subakan



FONDAZIONE BRUNO KESSLER
UNIVERSITÉ LAVAL
UNIVERSITÉ Concordia UNIVERSITY
Mila

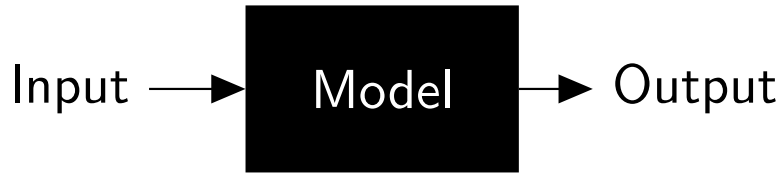
Explainable Machine Learning

- Black-box models

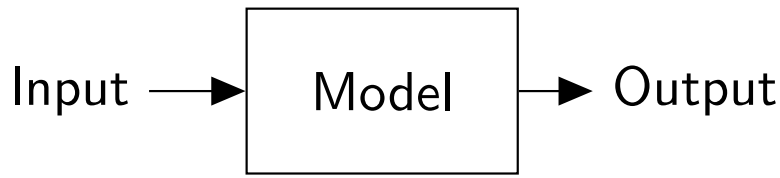


Explainable Machine Learning

- Black-box models

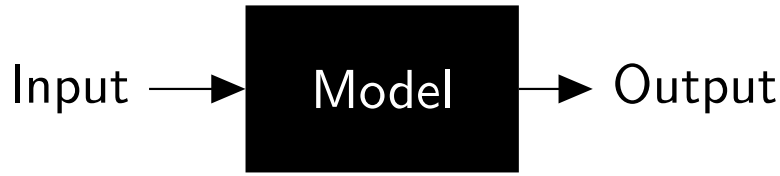


- Explainable Models

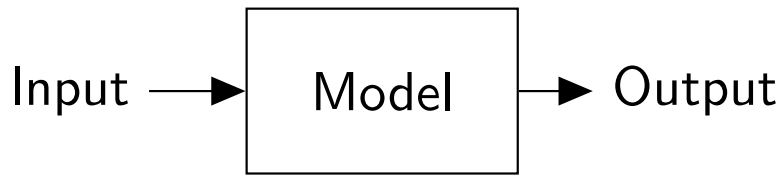


Explainable Machine Learning

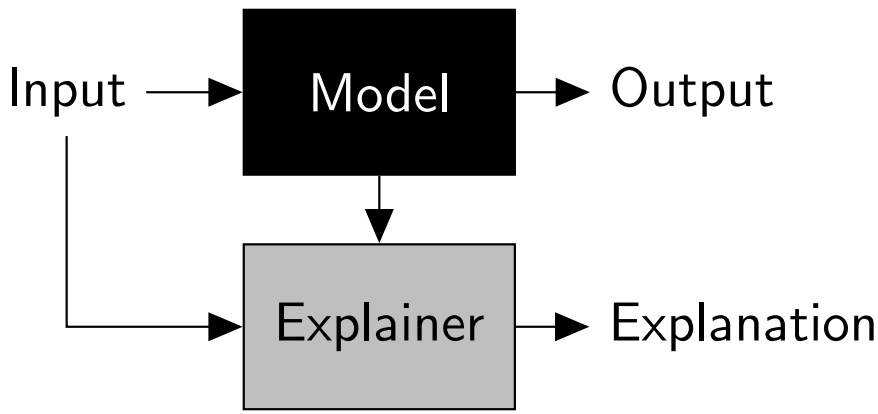
- Black-box models



- Explainable Models

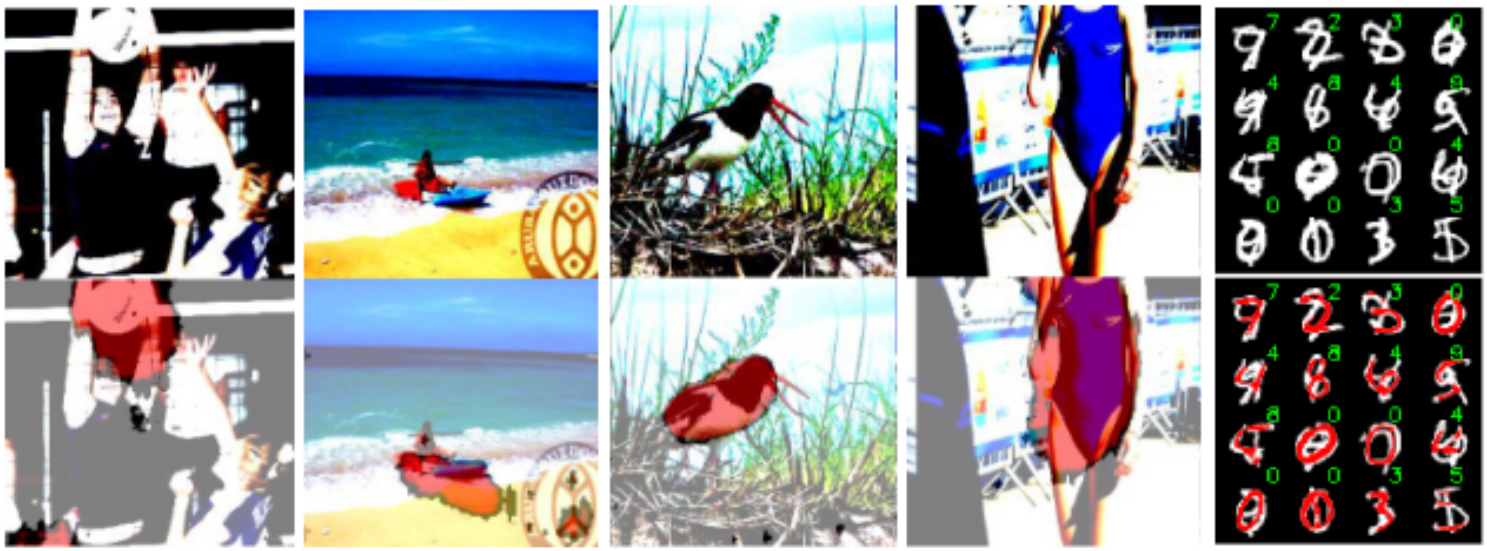


- Posthoc Explanations



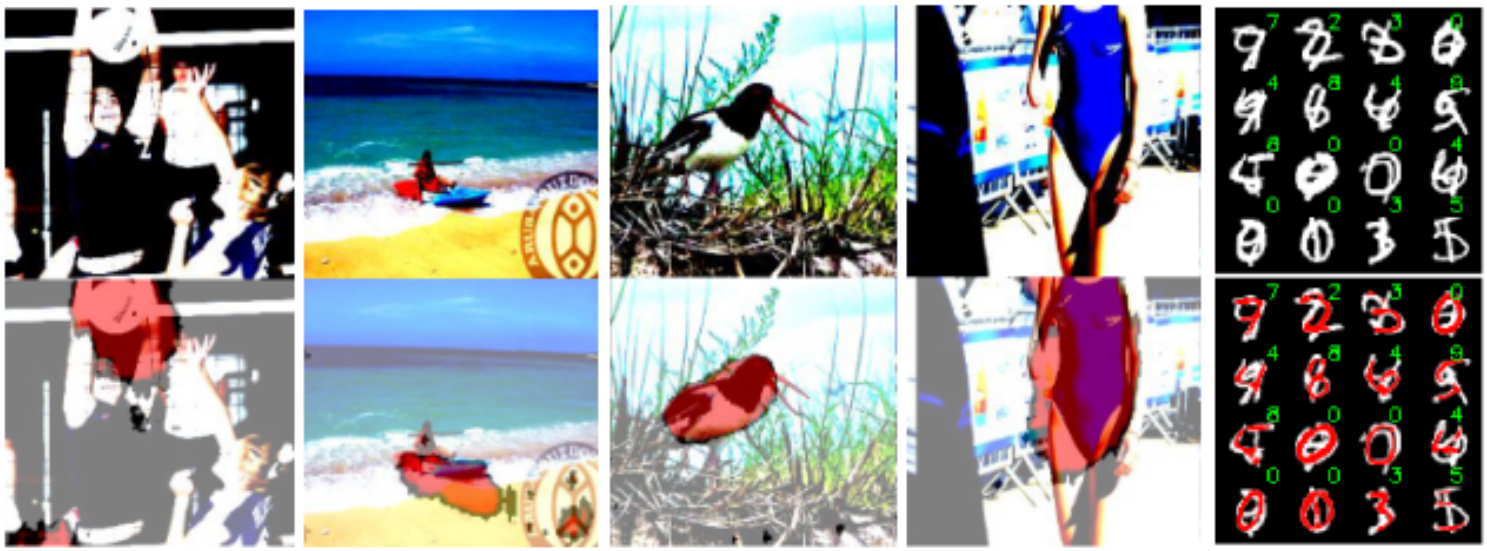
Explanations

- Saliency maps are commonly used in computer vision for producing explanations.



Explanations

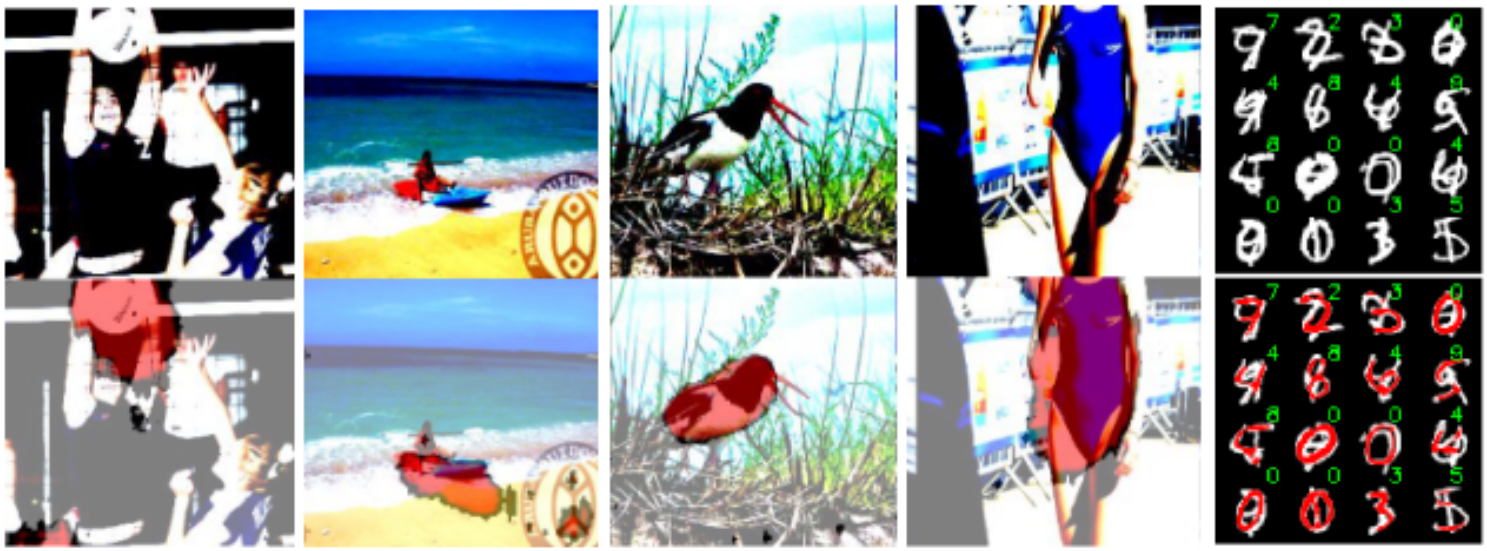
- Saliency maps are commonly used in computer vision for producing explanations.



- The explanations should **faithfully** follow the original model.

Explanations

- Saliency maps are commonly used in computer vision for producing explanations.

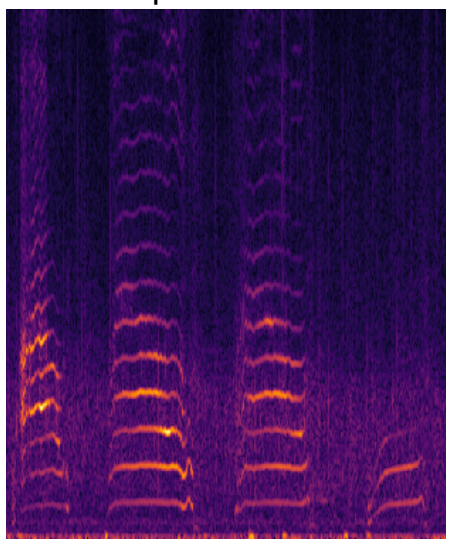


- The explanations should **faithfully** follow the original model.
- **Faithful** and **understandable** explanations are important for domains where decisions are critical!

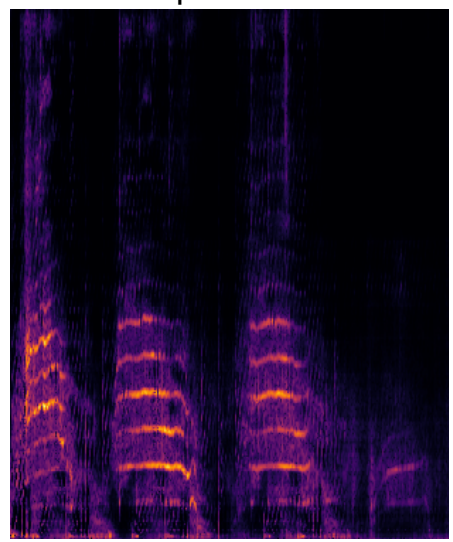
Contributions

- We develop an **understandable** and **faithful** (SOTA) posthoc explanation method for audio classifiers.

Input Audio



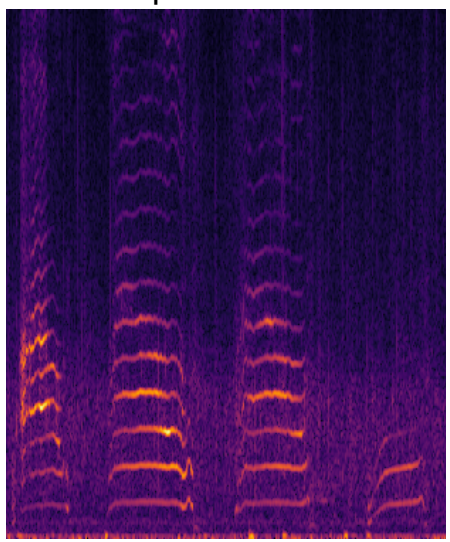
Explanation



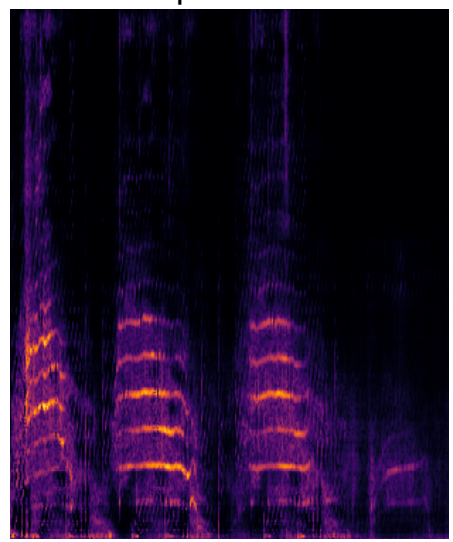
Contributions

- We develop an **understandable** and **faithful** (SOTA) posthoc explanation method for audio classifiers.

Input Audio



Explanation

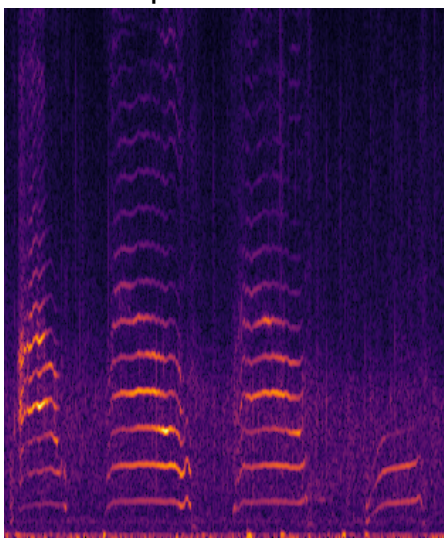


- Our method is agnostic to classifier input domain, and generates **listenable** explanations.

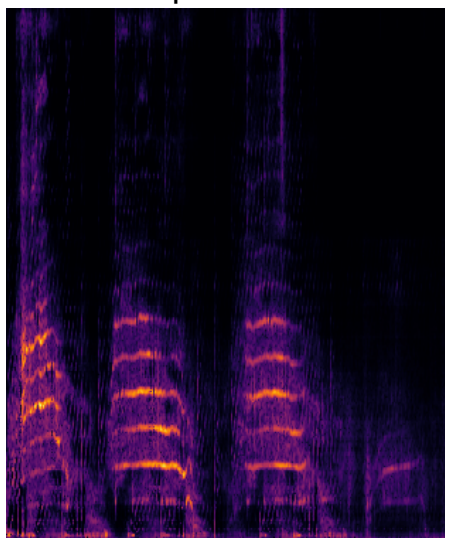
Contributions

- We develop an **understandable** and **faithful** (SOTA) posthoc explanation method for audio classifiers.

Input Audio



Explanation



- Our method is agnostic to classifier input domain, and generates **listenable** explanations.
- We propose a fine-tuning strategy that improves understandability/faithfulness trade-off.

Considerations

We would like to obtain

Considerations

We would like to obtain

- Faithful,

Considerations

We would like to obtain

- Faithful,
- Listenable,

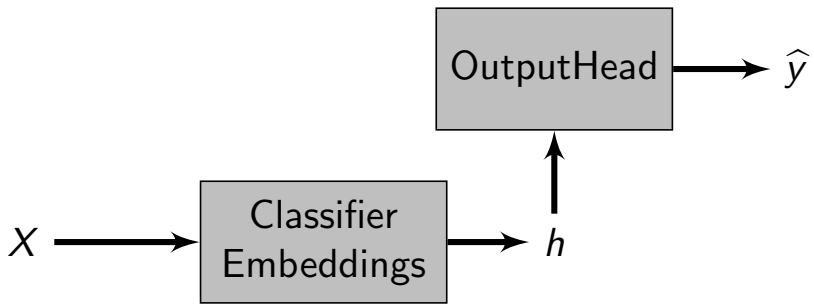
Considerations

We would like to obtain

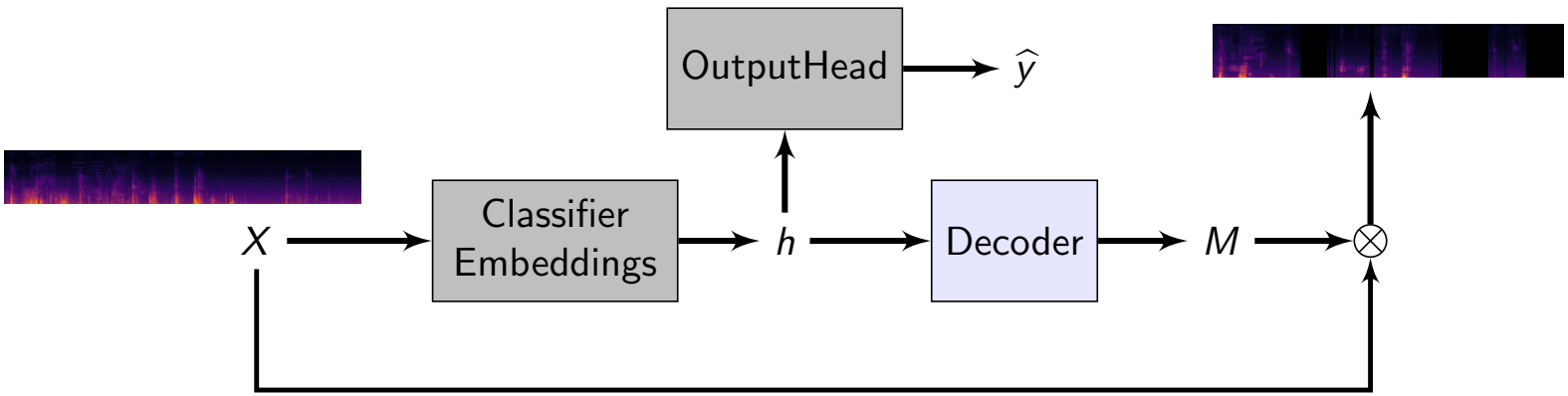
- Faithful,
- Listenable,
- Understandable

Posthoc Explanations for Audio Classifiers

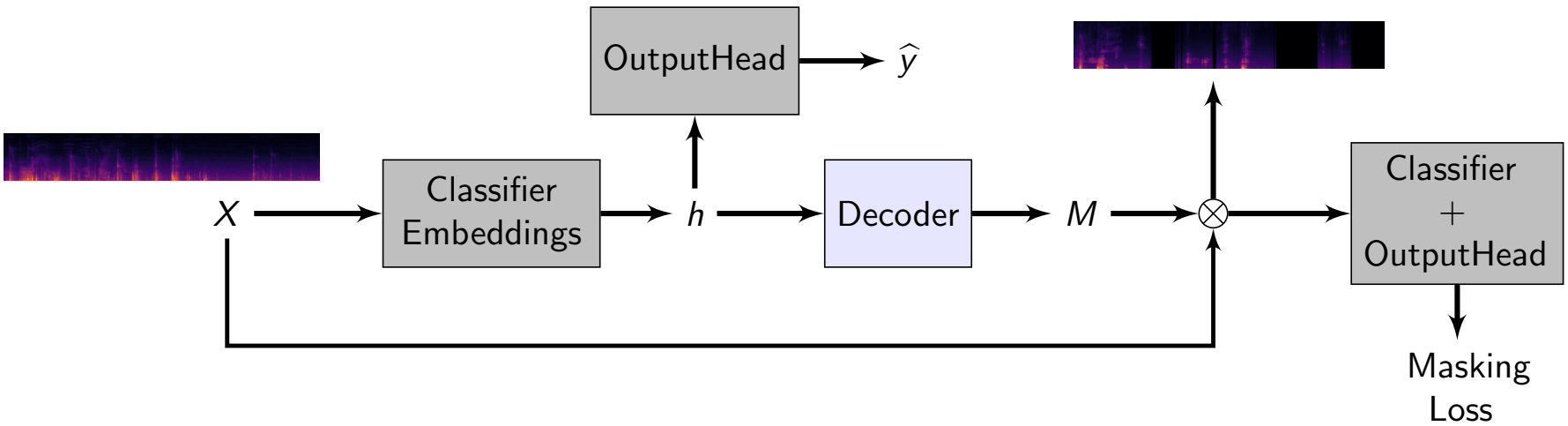
Listenable Maps for Audio Classifiers



Listenable Maps for Audio Classifiers

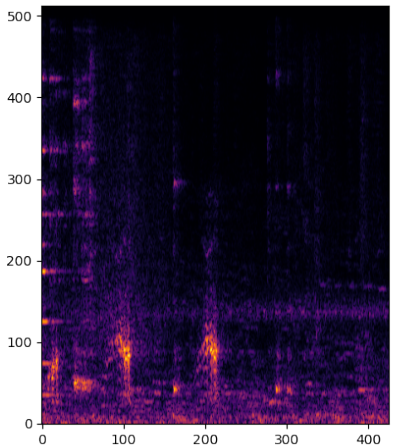


Listenable Maps for Audio Classifiers



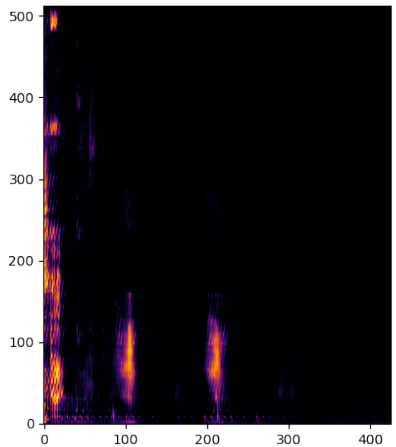
Optimization objective

Input Audio



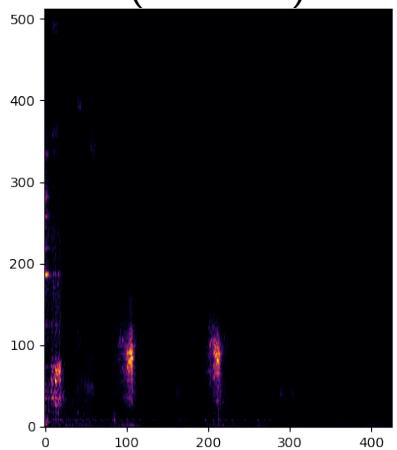
X

Saliency Map



$M_\theta(h)$

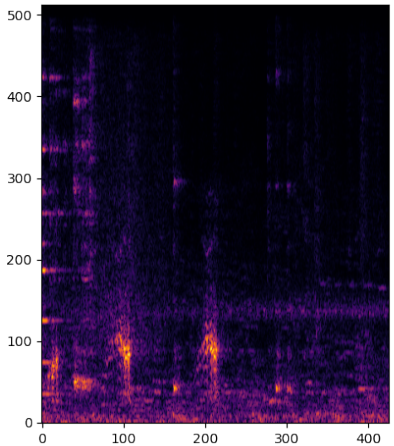
Explanation (Mask-in)



$M_\theta(h) \odot X$

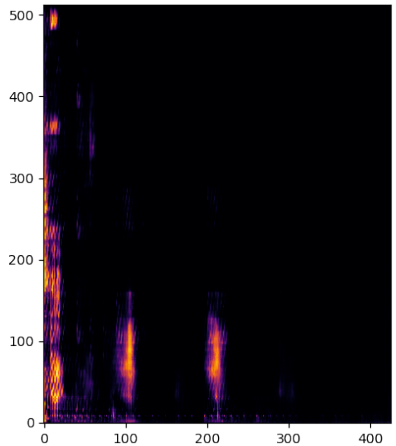
Optimization objective

Input Audio



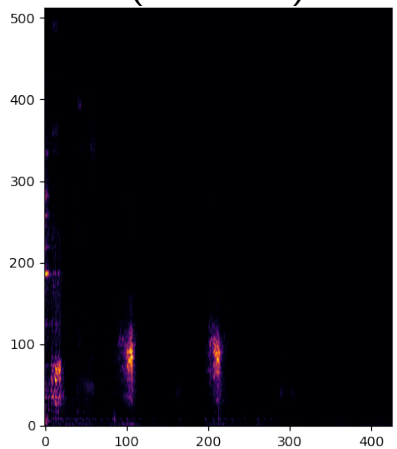
X

Saliency Map



$M_\theta(h)$

Explanation (Mask-in)

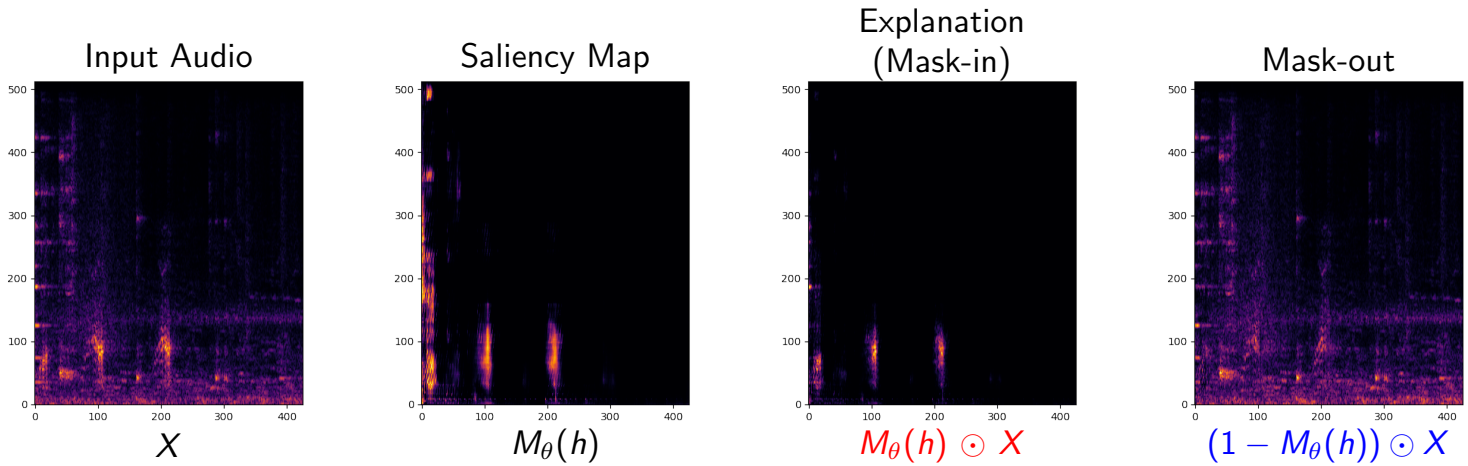


$M_\theta(h) \odot X$

$$\min_\theta \overbrace{\lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y})}^{\text{Mask-in}}$$

Maximizes the classifier agreement between the input and the explanation.

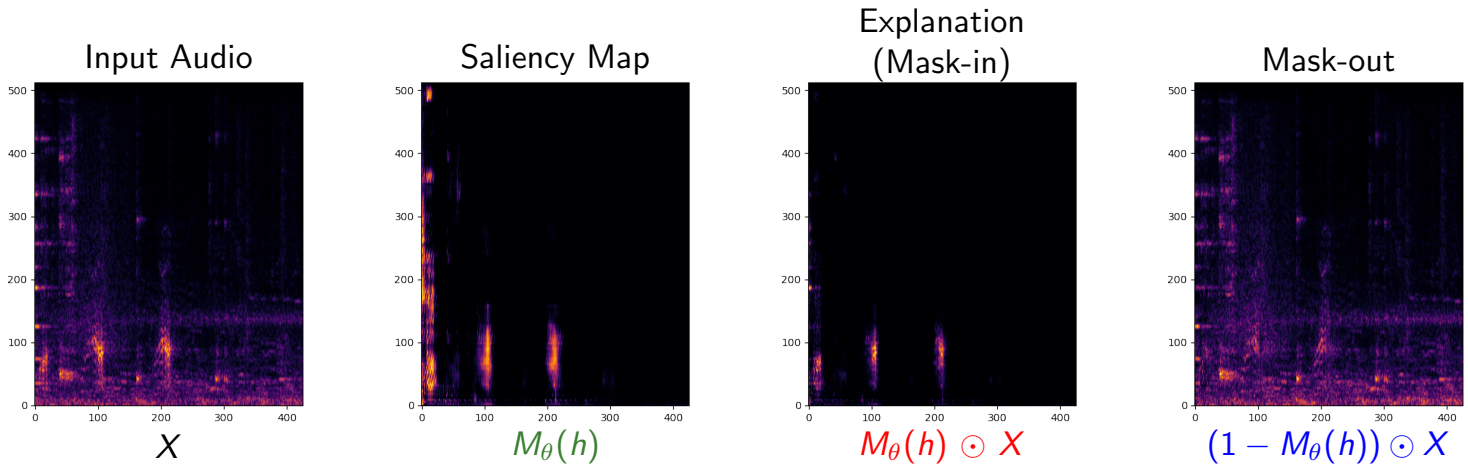
Optimization objective



$$\min_\theta \overbrace{\lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y})}^{\text{Mask-in}} \overbrace{- \lambda_{out}\mathcal{L}_{out}(\log f((1 - M_\theta(h)) \odot X), \widehat{y})}^{\text{Mask-out}}$$

Minimizes the classifier agreement of what is not in the explanation and the input.

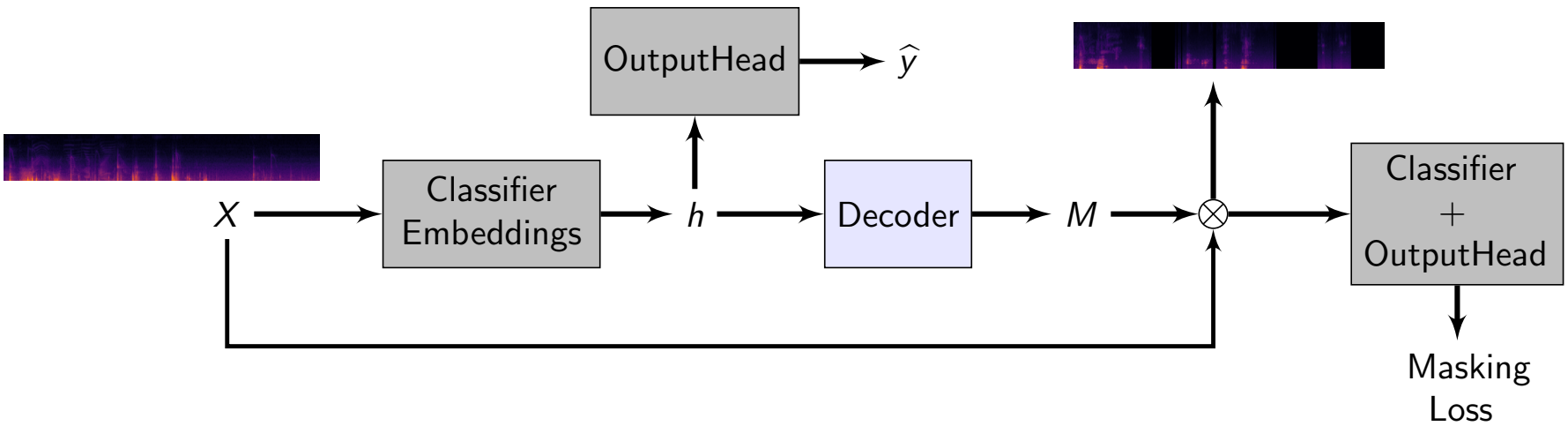
Optimization objective



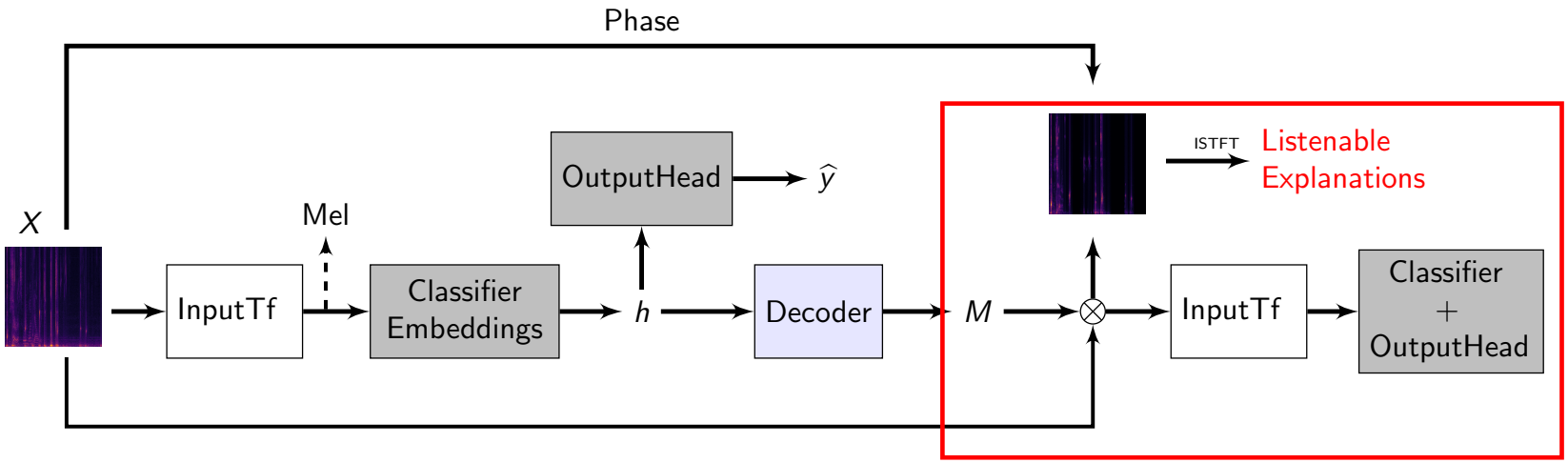
$$\min_\theta \overbrace{\lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y})}^{\text{Mask-in}} - \overbrace{\lambda_{out}\mathcal{L}_{out}(\log f((1 - M_\theta(h)) \odot X), \widehat{y})}^{\text{Mask-out}} + \overbrace{|M_\theta(h)|}^{\text{Mask Reg}}$$

Avoids trivial solutions.

Producing Listenable Explanations



Producing Listenable Explanations



$$\textit{Listenable Explanation} = \mathrm{ISTFT}\left((M_\theta(h) \odot X)e^{jX_{\mathrm{phase}}}\right)$$

Measuring faithfulness and understandability

- **Faithfulness**: Measures importance of explanations for classifier decisions
  - L2I-Faithfulness
  - Average-Increase
  - Average-Gain
  - Average-Drop
  - Input Fidelity
- **Structural metrics**: Measures the understandability of the explanations
  - Sparseness
  - Complexity

Understandability

$$\min_{\theta} \lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y})$$
$$-\lambda_{out}\mathcal{L}_{out}(\log f((1 - M_\theta(h)) \odot X), \widehat{y}) + \overbrace{|M_\theta(h)|}^{Regularizer}$$

Understandability

$$\min_{\theta} \lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y}) - \lambda_{out}\mathcal{L}_{out}(\log f((1 - M_\theta(h)) \odot X), \widehat{y})$$
$$+ \lambda_s \underbrace{\|M_\theta(h)\|_1}_{L_1} + \lambda_g \underbrace{\|M_\theta(h) \odot X - X_{clean}\|}_{Finetuning}$$

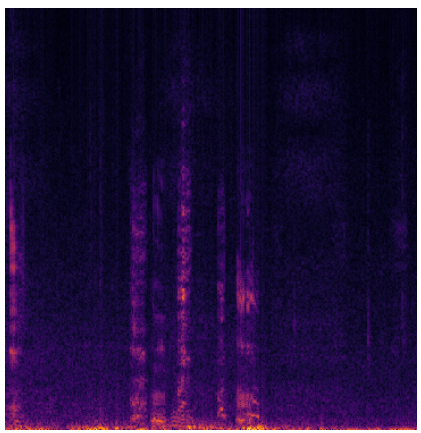
- L_1: Avoids trivial solutions (e.g. all 1s).
- *Finetuning*: Improves Understandability.
  - Used in a second stage, selectively.

Understandability

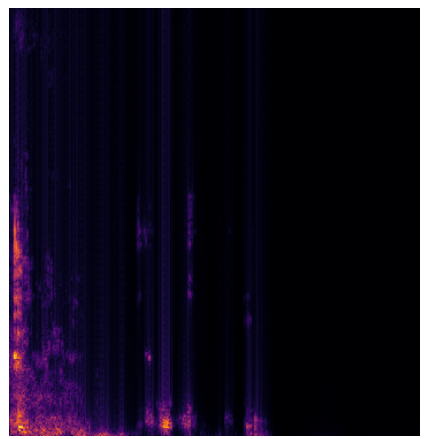
$$\min_{\theta} \lambda_{in}\mathcal{L}_{in}(\log f(M_\theta(h) \odot X), \widehat{y}) - \lambda_{out}\mathcal{L}_{out}(\log f((1 - M_\theta(h)) \odot X), \widehat{y})$$
$$+ \lambda_s \underbrace{\|M_\theta(h)\|_1}_{L_1} + \lambda_g \underbrace{\|M_\theta(h) \odot X - X_{clean}\|}_{Finetuning}$$

- L_1: Avoids trivial solutions (e.g. all 1s).
- *Finetuning*: Improves Understandability.
  - Used in a second stage, selectively.

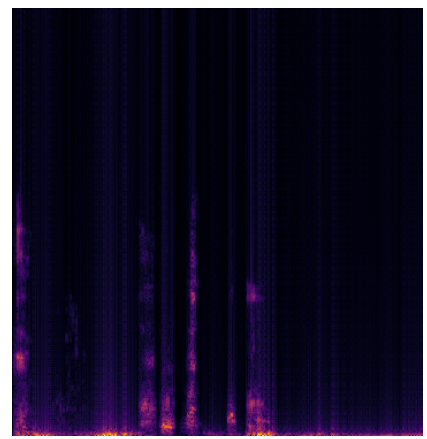
Input Audio



No finetuning



Finetuning



Experiments

- We produce explanations for classifiers trained on Sound Event Classification Datasets (**ESC50**, **US8k**).

Experiments

- We produce explanations for classifiers trained on Sound Event Classification Datasets (**ESC50**, **US8k**).
- We examine explanations on In-Domain (**ID**) and Out-of-Domain (**OOD**) cases.
  - ID: Plain datasets with data augmentation
  - OOD: Mixtures with different contaminating sources

Quantitative Results - ID

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.00	15.79	0.00	0.05	0.07	0.39	5.48
	Smoothgrad	0.00	15.71	0.00	0.03	0.05	0.42	5.32
	IG	0.25	15.45	0.01	0.07	0.13	0.43	5.11
	GradCAM	8.50	10.11	1.47	0.17	0.33	0.34	5.64
	Guided GradCAM	0.00	15.61	0.00	0.05	0.06	0.44	5.12
	Guided Backprop	0.00	15.66	0.00	0.05	0.06	0.39	5.47
	L2I, RT=0.2	1.63	12.78	0.42	0.11	0.15	0.25	5.50
	SHAP	0.00	15.79	0.00	0.05	0.06	0.43	5.24
	L-MAC (ours)	**36.25**	**1.15**	**23.50**	0.20	**0.42**	**0.47**	**4.71**
	L-MAC, FT, $\lambda_g = 4$ (ours)	32.37	1.98	18.74	**0.21**	0.41	0.43	5.20
Not Listenable (Mel)	Saliency	0.00	15.81	0.00	0.10	0.07	0.39	4.53
	Smoothgrad	0.00	15.61	0.00	0.07	0.04	0.39	4.54
	IG	0.00	15.55	0.00	0.12	0.08	0.42	4.36
	GradCAM	7.00	10.93	1.04	0.17	0.29	0.34	**4.72**
	Guided GradCAM	0.125	15.40	6.67	0.08	0.07	**0.45**	4.17
	Guided Backprop	0.125	15.54	0.00	0.10	0.08	0.39	4.53
	SHAP	0.00	15.57	0.00	0.11	0.08	0.41	4.42
	L-MAC (ours)	35.63	1.59	**24.28**	0.22	**0.42**	**0.45**	4.11
	L-MAC (ours) FT, $\lambda_g = 4$	**36.13**	**1.28**	21.15	**0.23**	**0.42**	0.32	4.71

Quantitative Results - ID

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.00	15.79	0.00	0.05	0.07	0.39	5.48
	Smoothgrad	0.00	15.71	0.00	0.03	0.05	0.42	5.32
	IG	0.25	15.45	0.01	0.07	0.13	0.43	5.11
	GradCAM	8.50	10.11	1.47	0.17	0.33	0.34	5.64
	Guided GradCAM	0.00	15.61	0.00	0.05	0.06	0.44	5.12
	Guided Backprop	0.00	15.66	0.00	0.05	0.06	0.39	5.47
	L2I, RT=0.2	1.63	12.78	0.42	0.11	0.15	0.25	5.50
	SHAP	0.00	15.79	0.00	0.05	0.06	0.43	5.24
	L-MAC (ours)	**36.25**	**1.15**	**23.50**	0.20	**0.42**	**0.47**	**4.71**
	L-MAC, FT, $\lambda_g = 4$ (ours)	32.37	1.98	18.74	**0.21**	0.41	0.43	5.20
Not Listenable (Mel)	Saliency	0.00	15.81	0.00	0.10	0.07	0.39	4.53
	Smoothgrad	0.00	15.61	0.00	0.07	0.04	0.39	4.54
	IG	0.00	15.55	0.00	0.12	0.08	0.42	4.36
	GradCAM	7.00	10.93	1.04	0.17	0.29	0.34	**4.72**
	Guided GradCAM	0.125	15.40	6.67	0.08	0.07	**0.45**	4.17
	Guided Backprop	0.125	15.54	0.00	0.10	0.08	0.39	4.53
	SHAP	0.00	15.57	0.00	0.11	0.08	0.41	4.42
	L-MAC (ours)	35.63	1.59	**24.28**	0.22	**0.42**	**0.45**	4.11
	L-MAC (ours) FT, $\lambda_g = 4$	**36.13**	**1.28**	21.15	**0.23**	**0.42**	0.32	4.71

Quantitative Results - ID

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.00	15.79	0.00	0.05	0.07	0.39	5.48
	Smoothgrad	0.00	15.71	0.00	0.03	0.05	0.42	5.32
	IG	0.25	15.45	0.01	0.07	0.13	0.43	5.11
	GradCAM	8.50	10.11	1.47	0.17	0.33	0.34	5.64
	Guided GradCAM	0.00	15.61	0.00	0.05	0.06	0.44	5.12
	Guided Backprop	0.00	15.66	0.00	0.05	0.06	0.39	5.47
	L2I, RT=0.2	1.63	12.78	0.42	0.11	0.15	0.25	5.50
	SHAP	0.00	15.79	0.00	0.05	0.06	0.43	5.24
	L-MAC (ours)	**36.25**	**1.15**	**23.50**	0.20	**0.42**	**0.47**	**4.71**
	L-MAC, FT, $\lambda_g = 4$ (ours)	32.37	1.98	18.74	**0.21**	0.41	0.43	5.20
Not Listenable (Mel)	Saliency	0.00	15.81	0.00	0.10	0.07	0.39	4.53
	Smoothgrad	0.00	15.61	0.00	0.07	0.04	0.39	4.54
	IG	0.00	15.55	0.00	0.12	0.08	0.42	4.36
	GradCAM	7.00	10.93	1.04	0.17	0.29	0.34	**4.72**
	Guided GradCAM	0.125	15.40	6.67	0.08	0.07	**0.45**	4.17
	Guided Backprop	0.125	15.54	0.00	0.10	0.08	0.39	4.53
	SHAP	0.00	15.57	0.00	0.11	0.08	0.41	4.42
	L-MAC (ours)	35.63	1.59	**24.28**	0.22	**0.42**	**0.45**	4.11
	L-MAC (ours) FT, $\lambda_g = 4$	**36.13**	**1.28**	21.15	**0.23**	**0.42**	0.32	4.71

- Finetuning does not harm faithfulness significantly.
- Generating listenable explanations does not decrease the alignment with the classifier.

Quantitative Results - ID

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.00	15.79	0.00	0.05	0.07	0.39	5.48
	Smoothgrad	0.00	15.71	0.00	0.03	0.05	0.42	5.32
	IG	0.25	15.45	0.01	0.07	0.13	0.43	5.11
	GradCAM	8.50	10.11	1.47	0.17	0.33	0.34	5.64
	Guided GradCAM	0.00	15.61	0.00	0.05	0.06	0.44	5.12
	Guided Backprop	0.00	15.66	0.00	0.05	0.06	0.39	5.47
	L2I, RT=0.2	1.63	12.78	0.42	0.11	0.15	0.25	5.50
	SHAP	0.00	15.79	0.00	0.05	0.06	0.43	5.24
	L-MAC (ours)	**36.25**	**1.15**	**23.50**	0.20	**0.42**	**0.47**	**4.71**
	L-MAC, FT, $\lambda_g = 4$ (ours)	32.37	1.98	18.74	**0.21**	0.41	0.43	5.20
Not Listenable (Mel)	Saliency	0.00	15.81	0.00	0.10	0.07	0.39	4.53
	Smoothgrad	0.00	15.61	0.00	0.07	0.04	0.39	4.54
	IG	0.00	15.55	0.00	0.12	0.08	0.42	4.36
	GradCAM	7.00	10.93	1.04	0.17	0.29	0.34	**4.72**
	Guided GradCAM	0.125	15.40	6.67	0.08	0.07	**0.45**	4.17
	Guided Backprop	0.125	15.54	0.00	0.10	0.08	0.39	4.53
	SHAP	0.00	15.57	0.00	0.11	0.08	0.41	4.42
	L-MAC (ours)	35.63	1.59	**24.28**	0.22	**0.42**	**0.45**	4.11
	L-MAC (ours) FT, $\lambda_g = 4$	**36.13**	**1.28**	21.15	**0.23**	**0.42**	0.32	4.71

- Finetuning does not harm faithfulness significantly.
- Generating listenable explanations does not decrease the alignment with the classifier.
- We have comparable structural metrics.

Quantitative Results - OOD (Audio Mixtures)

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.62	31.73	0.07	0.06	0.12	0.76	11.06
	Smoothgrad	0.12	31.84	0.00	0.06	0.13	0.83	10.66
	IG	0.37	31.15	0.03	0.12	0.26	0.87	10.22
	L2I	5.00	25.65	1.00	0.20	0.35	0.52	10.99
	GradCAM	14.12	17.62	7.46	0.25	0.00	0.91	9.66
	Guided GradCAM	0.00	31.74	0.00	0.07	0.11	0.89	10.24
	Guided Backprop	0.63	31.73	0.07	0.06	0.11	0.76	11.06
	SHAP	0.00	31.81	0.00	0.07	0.14	0.84	10.58
	L-MAC (ours)	**60.63**	**4.82**	**35.85**	**0.39**	**0.81**	**0.94**	**9.61**
	L-MAC FT, $\lambda_g = 4$ (ours)	50.75	6.73	26.00	**0.39**	0.78	0.84	10.51
Not Listenable (Mel)	Saliency	0.38	31.64	0.01	0.15	0.12	0.77	9.17
	Smoothgrad	0.25	31.66	0.01	0.14	0.11	0.79	9.03
	IG	0.12	31.52	0.01	0.19	0.19	0.84	8.62
	GradCAM	19.88	18.85	4.67	0.34	0.69	0.66	9.49
	Guided GradCAM	0.00	31.68	0	0.14	0.12	0.89	10.24
	Guided Backprop	0.38	31.64	0.01	0.15	0.12	0.77	9.16
	SHAP	0.25	31.60	0.00	0.17	0.15	0.82	8.81
	L-MAC (ours)	60.25	**4.84**	**34.72**	**0.44**	0.80	**0.90**	**8.29**
	L-MAC - FT, $\lambda_g = 4$ (ours)	**60.75**	**4.84**	29.34	**0.44**	**0.83**	0.64	9.38

Quantitative Results - OOD (Audio Mixtures)

	Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
Listenable (STFT→Mel)	Saliency	0.62	31.73	0.07	0.06	0.12	0.76	11.06
	Smoothgrad	0.12	31.84	0.00	0.06	0.13	0.83	10.66
	IG	0.37	31.15	0.03	0.12	0.26	0.87	10.22
	L2I	5.00	25.65	1.00	0.20	0.35	0.52	10.99
	GradCAM	14.12	17.62	7.46	0.25	0.00	0.91	9.66
	Guided GradCAM	0.00	31.74	0.00	0.07	0.11	0.89	10.24
	Guided Backprop	0.63	31.73	0.07	0.06	0.11	0.76	11.06
	SHAP	0.00	31.81	0.00	0.07	0.14	0.84	10.58
	L-MAC (ours)	**60.63**	**4.82**	**35.85**	**0.39**	**0.81**	**0.94**	**9.61**
	L-MAC FT, $\lambda_g = 4$ (ours)	50.75	6.73	26.00	**0.39**	0.78	0.84	10.51
Not Listenable (Mel)	Saliency	0.38	31.64	0.01	0.15	0.12	0.77	9.17
	Smoothgrad	0.25	31.66	0.01	0.14	0.11	0.79	9.03
	IG	0.12	31.52	0.01	0.19	0.19	0.84	8.62
	GradCAM	19.88	18.85	4.67	0.34	0.69	0.66	9.49
	Guided GradCAM	0.00	31.68	0	0.14	0.12	0.89	10.24
	Guided Backprop	0.38	31.64	0.01	0.15	0.12	0.77	9.16
	SHAP	0.25	31.60	0.00	0.17	0.15	0.82	8.81
	L-MAC (ours)	60.25	**4.84**	**34.72**	**0.44**	0.80	**0.90**	**8.29**
	L-MAC - FT, $\lambda_g = 4$ (ours)	**60.75**	**4.84**	29.34	**0.44**	**0.83**	0.64	9.38

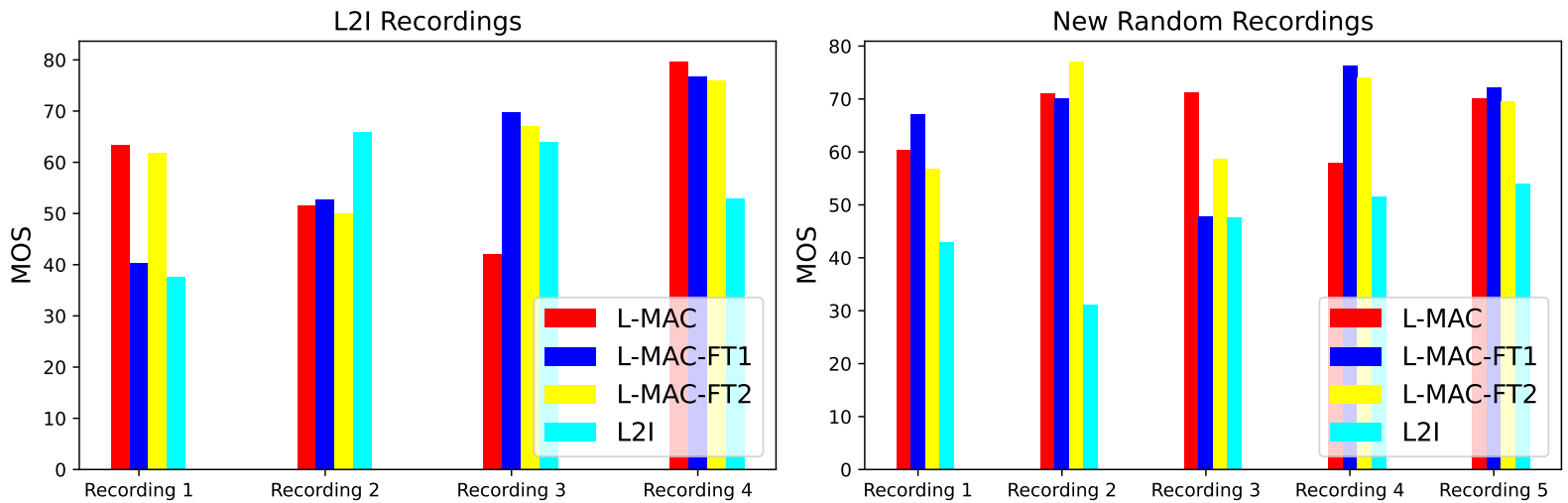
We observe the same outcome on US8k as well! (See the appendix)

User Study

1. How well does the explanation correspond to the part of the input audio associated with the given class?
2. While evaluating, please pay attention to audio quality also.

User Study

1. How well does the explanation correspond to the part of the input audio associated with the given class?
2. While evaluating, please pay attention to audio quality also.



- **Recording 1:**
  - **L-MAC**
  - **L2I [NeurIPS'22]**
- **OOD (Speech):**
  - **L-MAC**
  - **L2I [NeurIPS'22]**

Conclusions

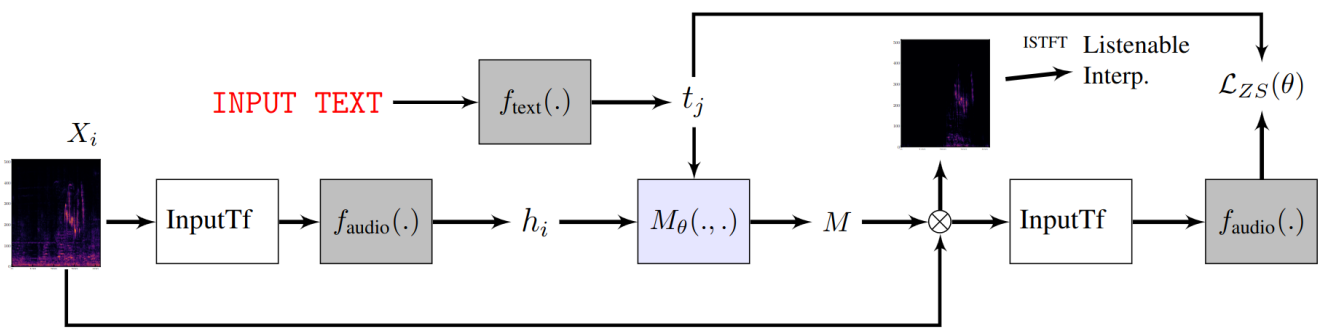
- We proposed a SOTA **posthoc explanation** method for audio classifers.
- Our method is agnostic to classifier input representation.
- Our method provides **understandable**, **listenable** and **faithful** explanations both in ID and OOD cases.
- Our code is available in SpeechBrain.

Conclusions

- We proposed a SOTA **posthoc explanation** method for audio classifers.
- Our method is agnostic to classifier input representation.
- Our method provides **understandable**, **listenable** and **faithful** explanations both in ID and OOD cases.
- Our code is available in SpeechBrain.

SpeechBrain

Follow-up Work: LMAC ZS: Listenable Maps for Zero-Shot Classifiers

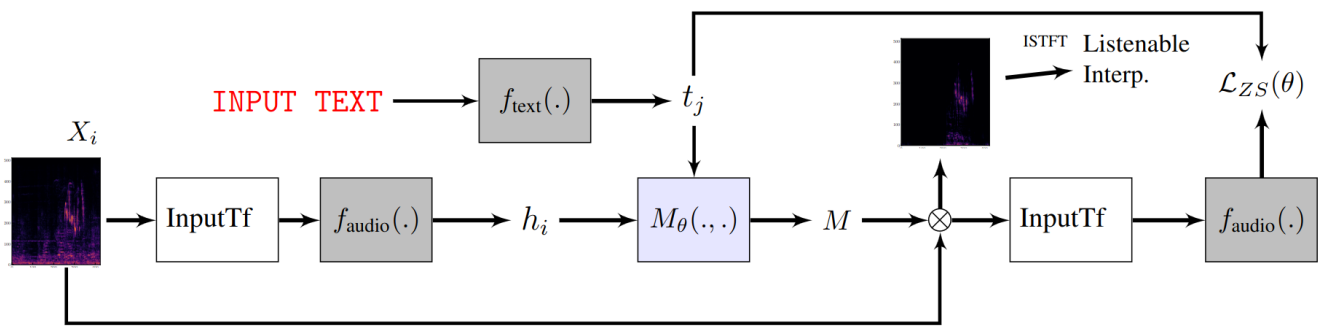


Thanks! Questions?

- We proposed a SOTA **posthoc explanation** method for audio classifers.
- Our method is agnostic to classifier input representation.
- Our method provides **understandable**, **listenable** and **faithful** explanations both in ID and OOD cases.
- Our code is available in SpeechBrain.

SpeechBrain

Follow-up Work: LMAC ZS: Listenable Maps for Zero-Shot Classifiers



Appendix A: More numbers

Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)	MM
	Classification on ESC50, White Noise Contamination, 38.6% accuracy							
Saliency	0.25	26.31	0.02	0.05	0.06	0.79	10.92	0.016
Smoothgrad	0.00	26.37	0.00	0.04	0.09	0.84	10.62	0.01
IG	0.75	25.60	0.56	0.10	0.21	0.82	10.65	0.01
L2I @ 0.2	0.00	19.41	0.21	0.11	0.04	**36.62**	**7.32**	0.12
GradCAM	8.87	20.88	1.24	0.28	0.51	0.69	11.25	0.18
Guided GradCAM	0.50	26.23	0.05	0.07	0.11	0.91	10.14	0.01
Guided Backprop	0.25	26.30	0.02	0.05	0.07	0.79	10.92	0.02
SHAP	0.12	26.34	0.001	0.05	0.12	0.86	10.40	0.004
L-MAC (ours)	**83.62**	**1.50**	**56.12**	**0.33**	**0.86**	0.92	10.03	0.06
All-ones baseline	0	0	0	0.34	1	N.A.	N.A.	1
	Classification on ESC50, LJSpeech Contamination, 79.3% accuracy							
Saliency	0.87	26.00	0.20	0.06	0.11	0.75	11.10	0.02
Smoothgrad	0.50	26.14	0.11	0.05	0.13	0.79	10.91	0.08
IG	0.37	25.70	0.01	0.11	0.25	0.87	10.14	0.00
L2I @ 0.2	1.75	29.49	0.27	0.15	0.18	0.79	**9.56**	0.16
GradCAM	20.37	13.49	2.63	0.28	0.73	0.66	11.33	0.22
Guided GradCAM	0.25	26.10	0.09	0.06	0.11	0.88	10.30	0.01
Guided Backprop	0.87	26.01	0.20	0.05	0.11	0.75	11.10	0.02
SHAP	0.00	26.14	0.00	0.06	0.16	0.79	10.81	0.01
L-MAC (ours)	**70.75**	**2.73**	**39.64**	**0.33**	**0.83**	**0.93**	9.70	0.05
All-ones baseline	0	0	0	0.35	1	N/A	N/A	1

Appendix B: Even more numbers

Metric	AI (↑)	AD (↓)	AG (↑)	FF (↑)	Fid-In (↑)	SPS (↑)	COMP (↓)
	Classification on ESC50, White Noise Contamination, 38.6% accuracy						
Saliency	0.25	26.31	0.02	0.05	0.06	0.79	10.92
Smoothgrad	0.00	26.37	0.00	0.04	0.09	0.84	10.62
IG	0.75	25.60	0.56	0.10	0.21	0.82	10.65
L2I @ 0.2	0.00	19.41	0.21	0.11	0.04	**36.62**	**7.32**
GradCAM	8.87	20.88	1.24	0.28	0.51	0.69	11.25
Guided GradCAM	0.50	26.23	0.05	0.07	0.11	0.91	10.14
Guided Backprop	0.25	26.30	0.02	0.05	0.07	0.79	10.92
SHAP	0.12	26.34	0.001	0.05	0.12	0.86	10.40
L-MAC (ours)	**83.62**	**1.50**	**56.12**	**0.33**	**0.86**	0.92	10.03
	Classification on ESC50, LJSpeech Contamination, 79.3% accuracy						
Saliency	0.87	26.00	0.20	0.06	0.11	0.75	11.10
Smoothgrad	0.50	26.14	0.11	0.05	0.13	0.79	10.91
IG	0.37	25.70	0.01	0.11	0.25	0.87	10.14
L2I @ 0.2	1.75	29.49	0.27	0.15	0.18	0.79	**9.56**
GradCAM	20.37	13.49	2.63	0.28	0.73	0.66	11.33
Guided GradCAM	0.25	26.10	0.09	0.06	0.11	0.88	10.30
Guided Backprop	0.87	26.01	0.20	0.05	0.11	0.75	11.10
SHAP	0.00	26.14	0.00	0.06	0.16	0.79	10.81
L-MAC (ours)	**70.75**	**2.73**	**39.64**	**0.33**	**0.83**	**0.93**	9.70

- **Speech 1:**
  - ▶ **L-MAC**
- **Speech 2:**
  - ▶ **L-MAC**
- **WN 1:**
  - ▶ **L-MAC**
- **WN 2:**
  - ▶ **L-MAC**