

Energy-based Multi-Modal Attention

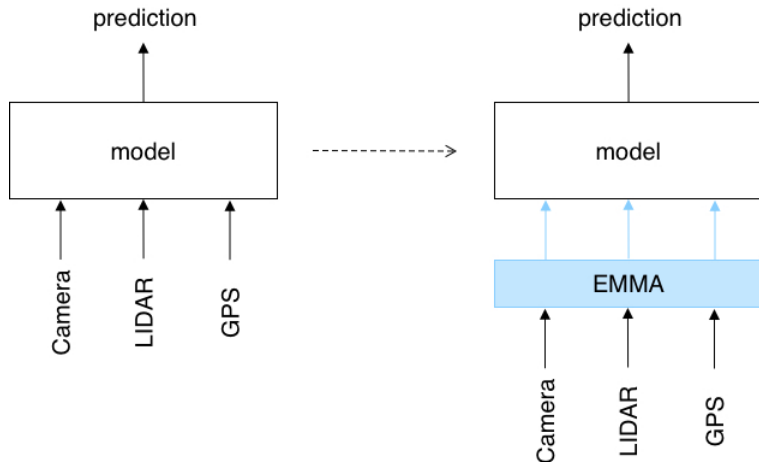
EMMA

Aurélien Werenne
Supervisor: Dr. Raphaël Marée



September 10, 2019

Problem & Solution



Module learns how to compute attention scores based on the three properties,

Module learns how to compute attention scores based on the three properties,

- Relevance: intrinsic informativeness for the predictive task at hand.

Module learns how to compute attention scores based on the three properties,

- Relevance: intrinsic informativeness for the predictive task at hand.
- Failure intensity: the propensity to trigger undesirable activations in the neural network.

Module learns how to compute attention scores based on the three properties,

- Relevance: intrinsic informativeness for the predictive task at hand.
- Failure intensity: the propensity to trigger undesirable activations in the neural network.
- Coupling: the interdependencies with the other modes.

Generalize robustness by using

- Capacity regularizer: minimize total amount of distributed attention.
- Energy regularizer: maximize influence of failure intensity.

	Hyperparameters			F1-score		
	ρ	λ_e	λ_c	uncorrupted	IP noisy	DM noisy
IP-only				0.8235	0.5926	
DM-only				0.6612		0.3920
base				0.8830	0.6441	0.6569
without				0.8671	0.7097	0.7683
other				0.7726	0.6129	0.6882
with	10^{-4}	10^{-3}	10^{-2}	0.8881	0.7333	0.8077
with	10^{-4}	0	10^{-2}	0.8849	0.7285	0.8183
with	10^{-4}	10^{-4}	10^{-2}	0.8945	0.7333	0.8182
with	10^{-3}	10^{-3}	0	0.8809	0.7347	0.8186
with	10^{-4}	10^{-2}	10^{-3}	0.8736	0.7383	0.7848
with	10^{-1}	10^{-2}	0	0.8826	0.7467	0.7925
with	10^{-4}	10^{-3}	0	0.8786	0.7190	0.7826
with	10^{-3}	10^{-1}	10^{-2}	0.8800	0.7432	0.8344
with	10^{-4}	0	10^{-4}	0.8723	0.7051	0.7853
with	10^{-4}	10^{-4}	10^{-3}	0.8794	0.7053	0.7853

Robustness generalization

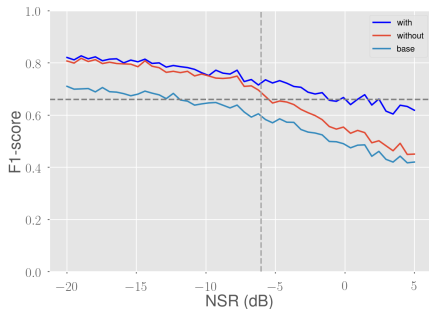


Figure 1: F1-score, noisy IP-mode

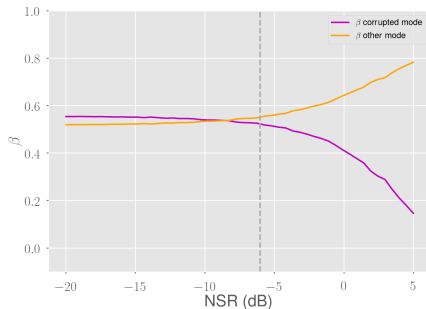


Figure 2: attention score, noisy IP-mode

Robustness generalization for DM mode

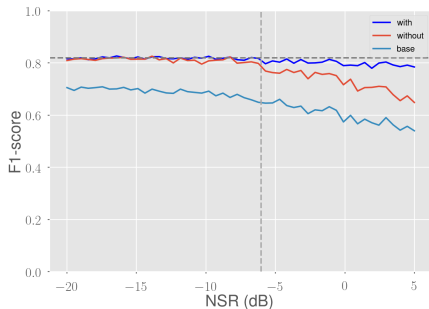


Figure 3: F1-score, noisy DM-mode

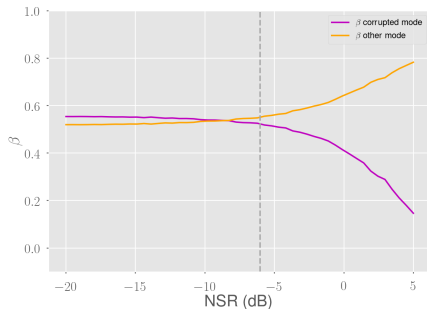


Figure 4: attention score, noisy DM-mode

Influence of capacity minimization

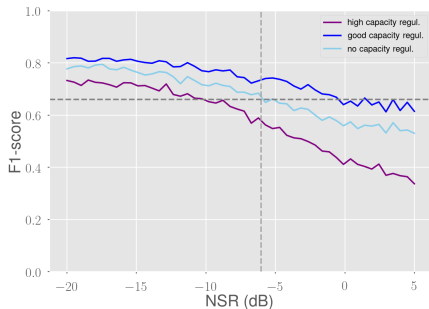


Figure 5: F1-score, noisy IP-mode

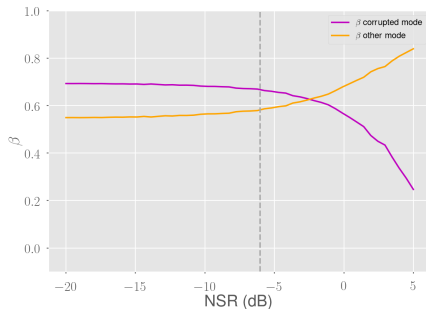


Figure 6: No capacity regularization

Influence of capacity minimization

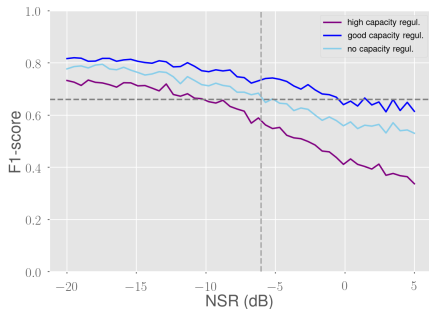


Figure 5: F1-score, noisy IP-mode

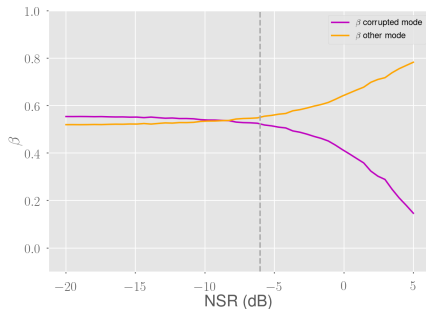


Figure 7: Good capacity regularization

Influence of capacity minimization

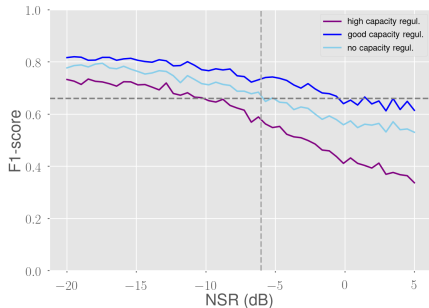


Figure 5: F1-score, noisy IP-mode

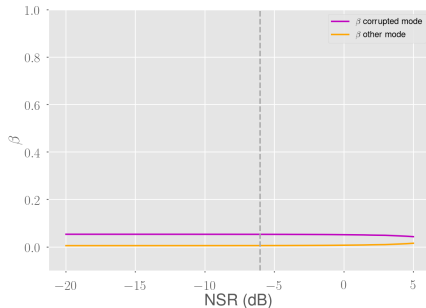


Figure 8: High capacity regularization

- **Problem?** Failing mode(s) \Rightarrow bad predictions.
- **Hypothesis?** Some redundancy of information between the modes.
- **Solution?** Attention module & regularizers.

Backup slides

Influence of energy regularizer

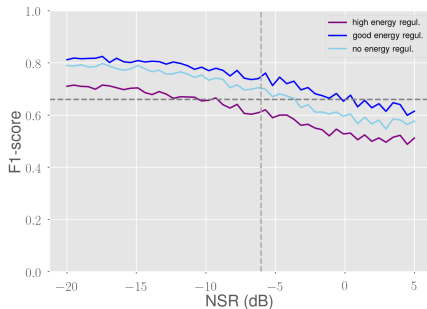


Figure 9: F1-score, noisy IP-mode

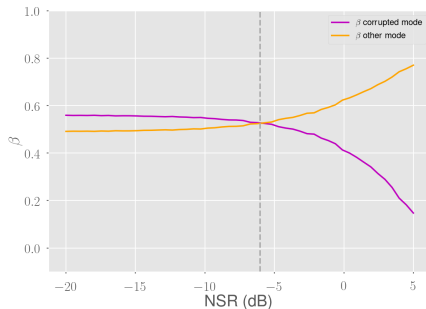


Figure 10: No energy regularization

Influence of energy regularizer

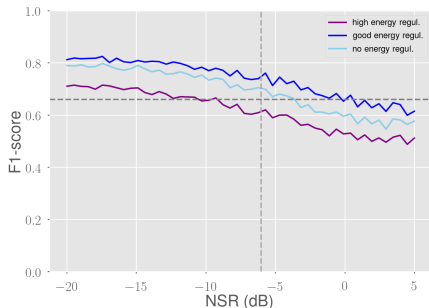


Figure 9: F1-score, noisy IP-mode

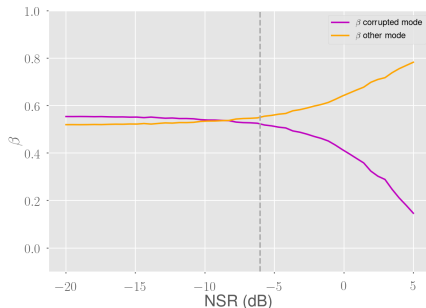


Figure 11: Good energy regularization

Influence of energy regularizer

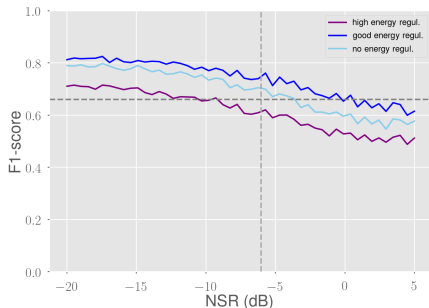


Figure 9: F1-score, noisy IP-mode

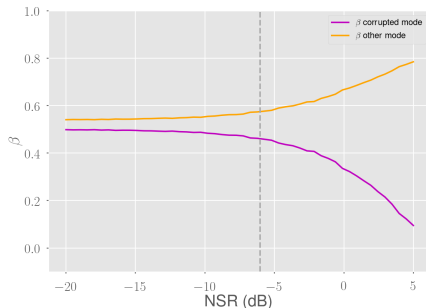


Figure 12: High energy regularization

Influence of energy regularizer for DM mode

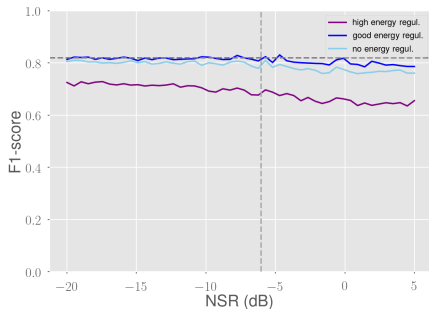


Figure 13: F1-score, noisy IP-mode

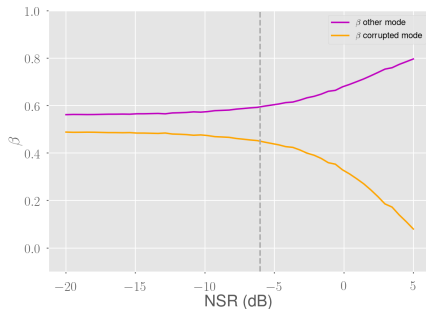


Figure 14: No energy regularization

Influence of energy regularizer for DM mode

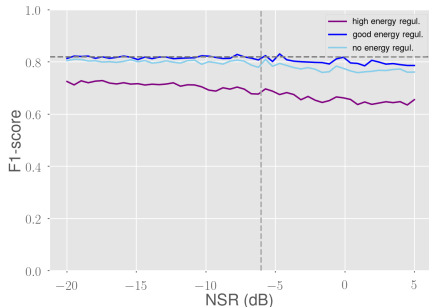


Figure 13: F1-score, noisy IP-mode

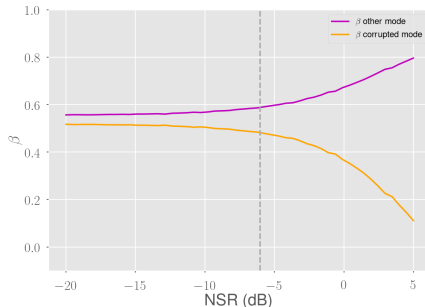


Figure 15: Good energy regularization

Influence of energy regularizer for DM mode

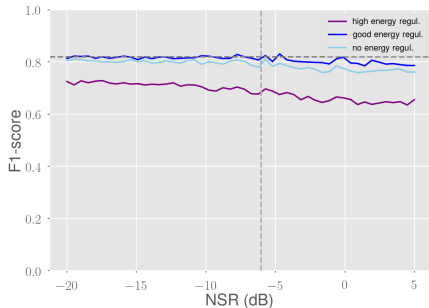


Figure 13: F1-score, noisy IP-mode

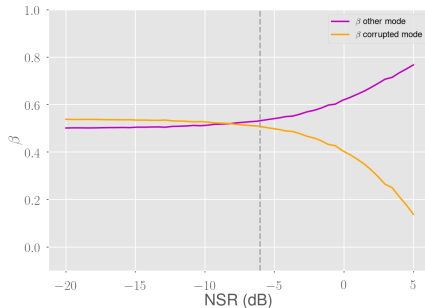


Figure 16: High energy regularization

Influence of capacity minimization for DM mode

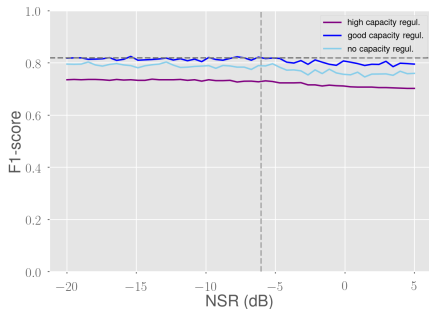


Figure 17: F1-score, noisy IP-mode

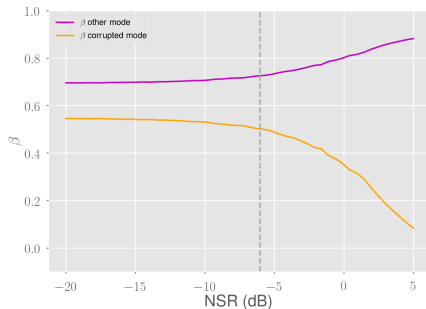


Figure 18: No capacity regularization

Influence of capacity minimization for DM mode

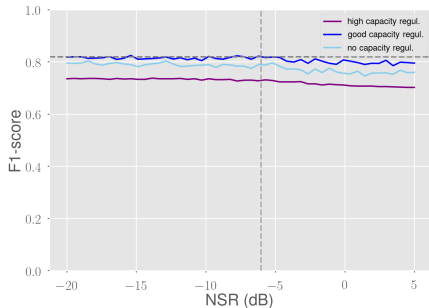


Figure 17: F1-score, noisy IP-mode

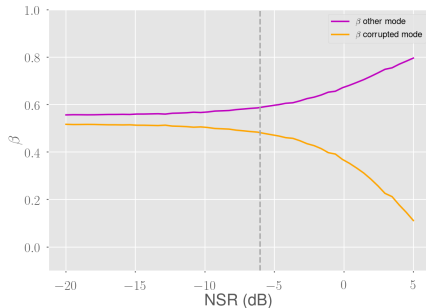


Figure 19: Good capacity regularization

Influence of capacity minimization for DM mode

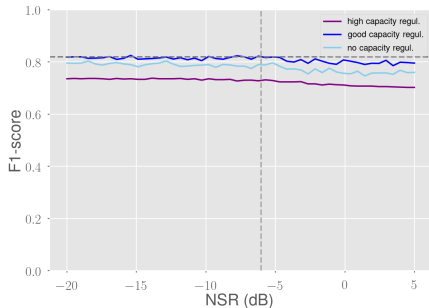


Figure 17: F1-score, noisy IP-mode

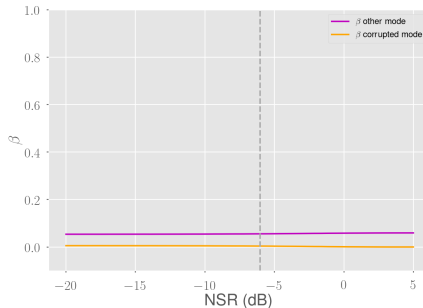


Figure 20: High capacity regularization

- 1 Failure intensity, Ψ_i
- 2 Self-energy, $e_i = w_i \Psi_i + b_i$
- 3 Shared energies, $e_{ij} = w_{ij} e_i^{\gamma_{ij}} e_j^{1-\gamma_{ij}}$
- 4 Modal energy, $E_i = e_i + \sum_{j \neq i} e_{ij}$
- 5 Importance score, $\alpha_i = \frac{1}{Z} e^{-\rho E_i}$
- 6 Attention score, $\beta_i = \max[0, \tanh(g_a \alpha_i - b_a)]$

$$\tilde{\mathcal{L}} = \mathcal{L}(y, \hat{y}) + \lambda_c(g_a - b_a) \quad (1)$$

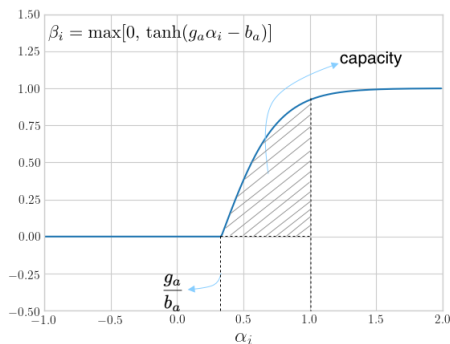


Figure 21: Attention function

$$\tilde{\mathcal{L}} = \mathcal{L}(y, \hat{y}) - \lambda_e \Omega \quad (2)$$

with

$$\Omega = \sum_{k=1}^M \xi_k \log(\alpha_k) \quad \text{and} \quad \xi_k = \begin{cases} \xi_- = -1 & \text{if } \mathbf{x}_k \text{ is corrupted} \\ \xi_+ = +1 & \text{otherwise} \end{cases} \quad (3)$$

If M is even,

$$E_i(\mathbf{x}_i; \boldsymbol{\theta}_i^{(0)} - \epsilon \lambda_e \rho \xi_i \nabla_{\boldsymbol{\theta}_i} E_i) \approx E_i(\mathbf{x}_i; \boldsymbol{\theta}_i^{(0)}) - \epsilon \lambda_e \rho \xi_i (\nabla_{\boldsymbol{\theta}_i} E_i)^T \nabla_{\boldsymbol{\theta}_i} E_i \quad (4)$$

If M is uneven,

$$\boldsymbol{\theta}_i \leftarrow \begin{cases} \boldsymbol{\theta}_i - \epsilon \lambda_e \rho (1 - \alpha_i) \nabla_{\boldsymbol{\theta}_i} E_i, & \text{if } i \text{ is uncorrupted} \\ \boldsymbol{\theta}_i + \epsilon \lambda_e \rho (1 + \alpha_i) \nabla_{\boldsymbol{\theta}_i} E_i & \text{otherwise} \end{cases} \quad (5)$$

Multiple noisy modes

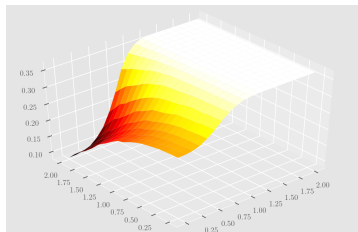


Figure 22: β -IP

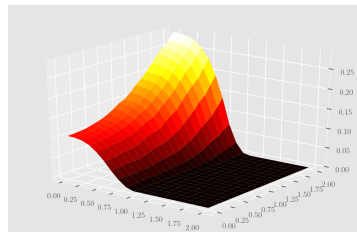


Figure 23: β -DM

$$\frac{\partial \Psi(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \propto -\frac{\partial \log p(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \not\Rightarrow \Psi \propto -\log p \quad (6)$$

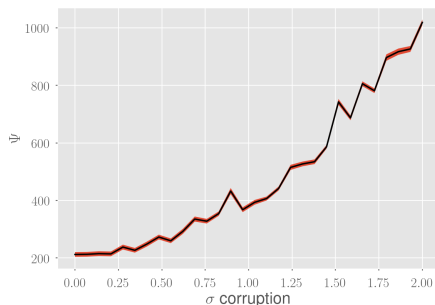


Figure 24: noisy DM-mode

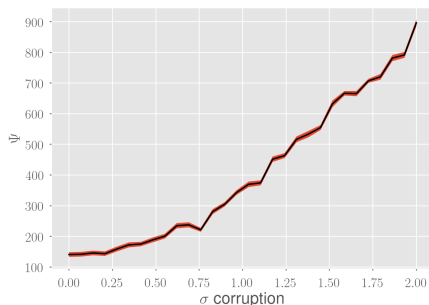


Figure 25: noisy IP-mode

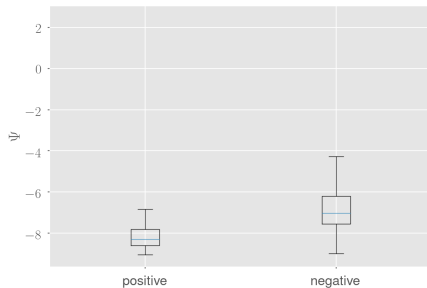


Figure 26: DM-mode

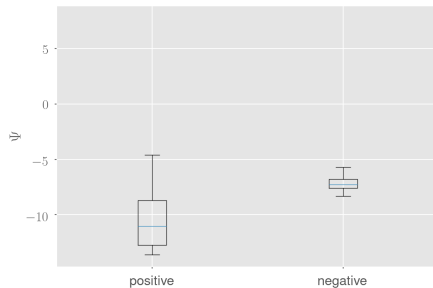


Figure 27: IP-mode

F1-score at NSR = 5 dB:

groundtruth	noise	Correct	Wrong
1	normal	0.9029	0.09701
1	IP-noisy	0.6825	0.3174
1	DM-noisy	0.7297	0.2702
0	normal	0.9843	0.01563
0	IP-noisy	0.8770	0.1229
0	DM-noisy	0.9503	0.04962

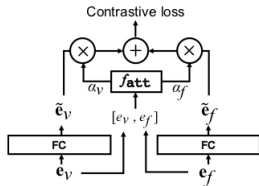


Figure 28: Model

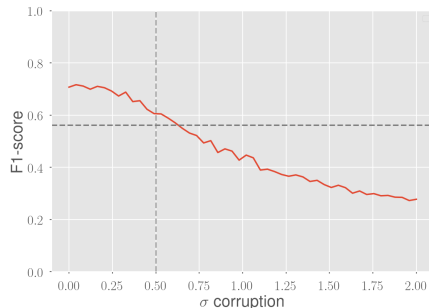


Figure 29: F1-score