Robustly overfitting latents for flexible neural image compression

Yura Perugachi Diaz
Arwin Gansekoele &
Sandjai Bhulai

Neural Image Compression

- State-of-the-art models based on variational autoencoders with encoder-decoder structure
- Models are trained to minimize the expected rate-distortion (R-D) costs:

$$\mathcal{L} = \mathcal{R} + \lambda \mathcal{D}$$

- **Aim:** find latent representation with best trade-off between length of bitstream & quality of reconstructed image

- In practice: limited capacity when it comes to optimization and generalization

Latent refinement

- Refining only the latents of pre-trained models:
  - Improved compression results per image without re-training model
- Variable y is **iteratively adapted** using differentiable operations test time
  - **Aim:** Find more optimal discrete latent representation $\hat{y}$

Minimization problem to be solved for image x:

$$\arg\min_{\hat{y}} \left[-\log_2 p_{\hat{y}}(\hat{y}) + \lambda d(x, \hat{x}) \right]$$

SGA

- In (Yang et al., 2020), propose **Stochastic Gumbel Annealing** (SGA) to optimize latents
- Soft-to-hard quantization method: quantizes continuous variable v into discrete representation for which gradients can be computed
- The logits are computed with atanh with following function to obtain **unnormalized log probabilities:**

$$\text{logits} = (\text{-atanh}(v_L)/\tau, -\text{atanh}(\mathrm{v}_R)/\tau)$$

- To obtain **probabilities**, a softmax is used over the logits, which gives probability $p(y)$
- Approximated by Gumbel-softmax distribution

SGA+

SGA

- Computation of *logits* is obtained with atanh
- Looking at probability space atanh:
    - **Sub-optimal**
    - Gradients tend to infinity when approaching limits of 0 and 1

SGA+

- Probabilities given by softmax over *logits* with function of choice
- Probabilities need to be:
    - Monotonic functions
    - Rounding down and up sums up to 1
- We extend SGA with SGA+
    - Contains **three methods** for *logits* computation
    - Overcomes issues SGA
    - Can be extended to **three class** rounding

Probability space

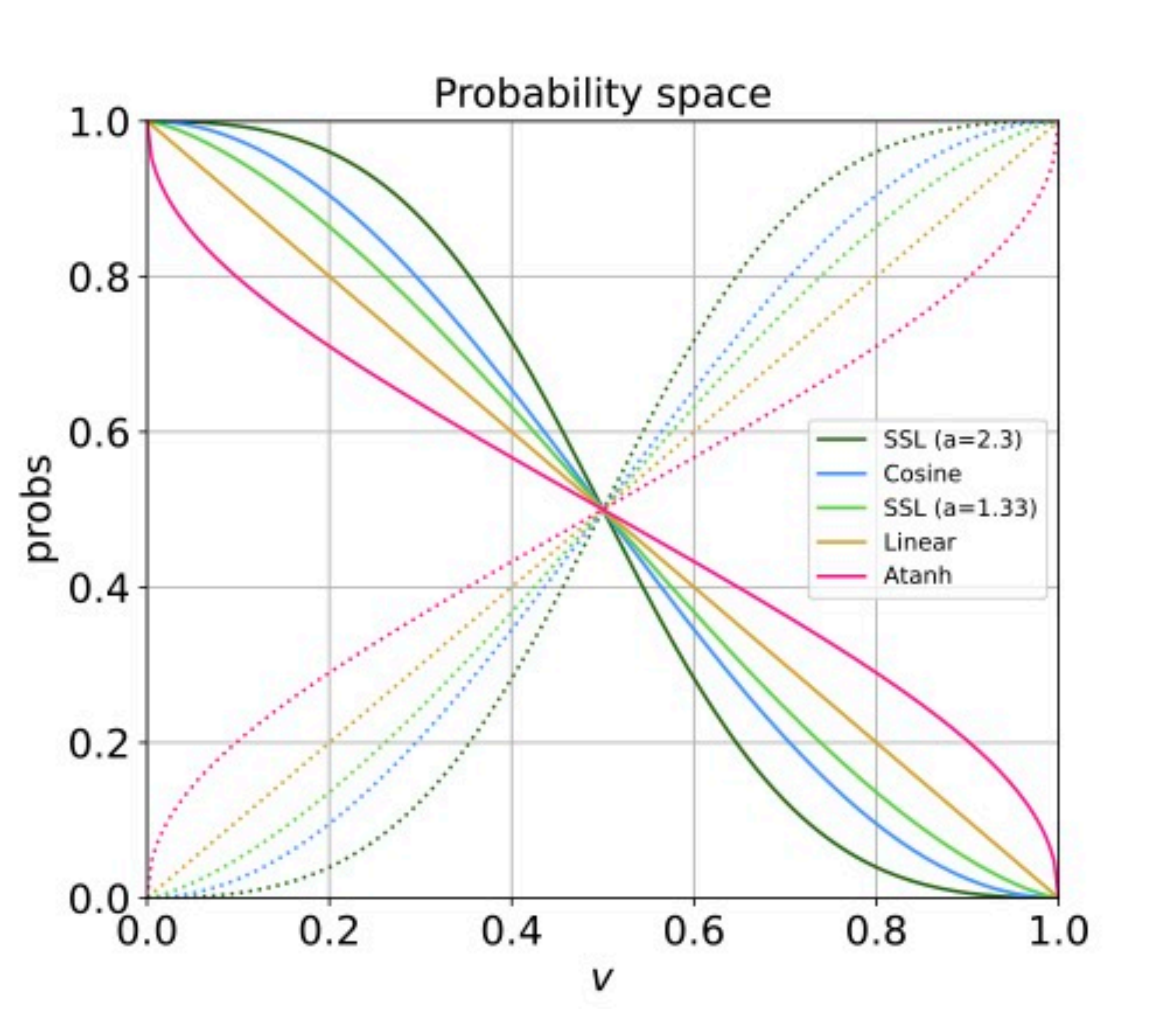

Figure 1: Probability space for several SGA+ methods

SGA+

Linear:

$$p(y = \lfloor v \rfloor) = 1 - (v - \lfloor v \rfloor)$$

- **Most natural case** since it linearly increases/decreases
- Prevents saturation or vanishing gradients completely
- Robust choice since constant gradients

Cosine:

$$p(y = \lfloor v \rfloor) = \cos^2\left(\frac{(v - \lfloor v \rfloor)\pi}{2}\right)$$

- Has low gradients in the area where atanh tends to have gradients that go to infinity for v close to corners

Sigmoid Scaled Logit (SSL):

$$p(y = \lfloor v \rfloor) = \sigma(-a\sigma^{-1}(v - \lfloor v \rfloor))$$

- Interpolates between all possible functions with its hyperparameter a

Experiments

Implementation details

- Two pre-trained hyperprior models to test SGA+ (Cheng et al., 2020)
- Package from **CompressAI**.
- Models were trained w.:

$\lambda = \{0.0016, 0.0032, 0.0075, 0.015, 0.03, 0.045\}$

- Using Kodak, Tecnick & CLIC dataset

Baseline methods

- Compare against **literature methods**:
  - Straight-Through Estimator
  - Uniform noise
  - SGA

Overall performance

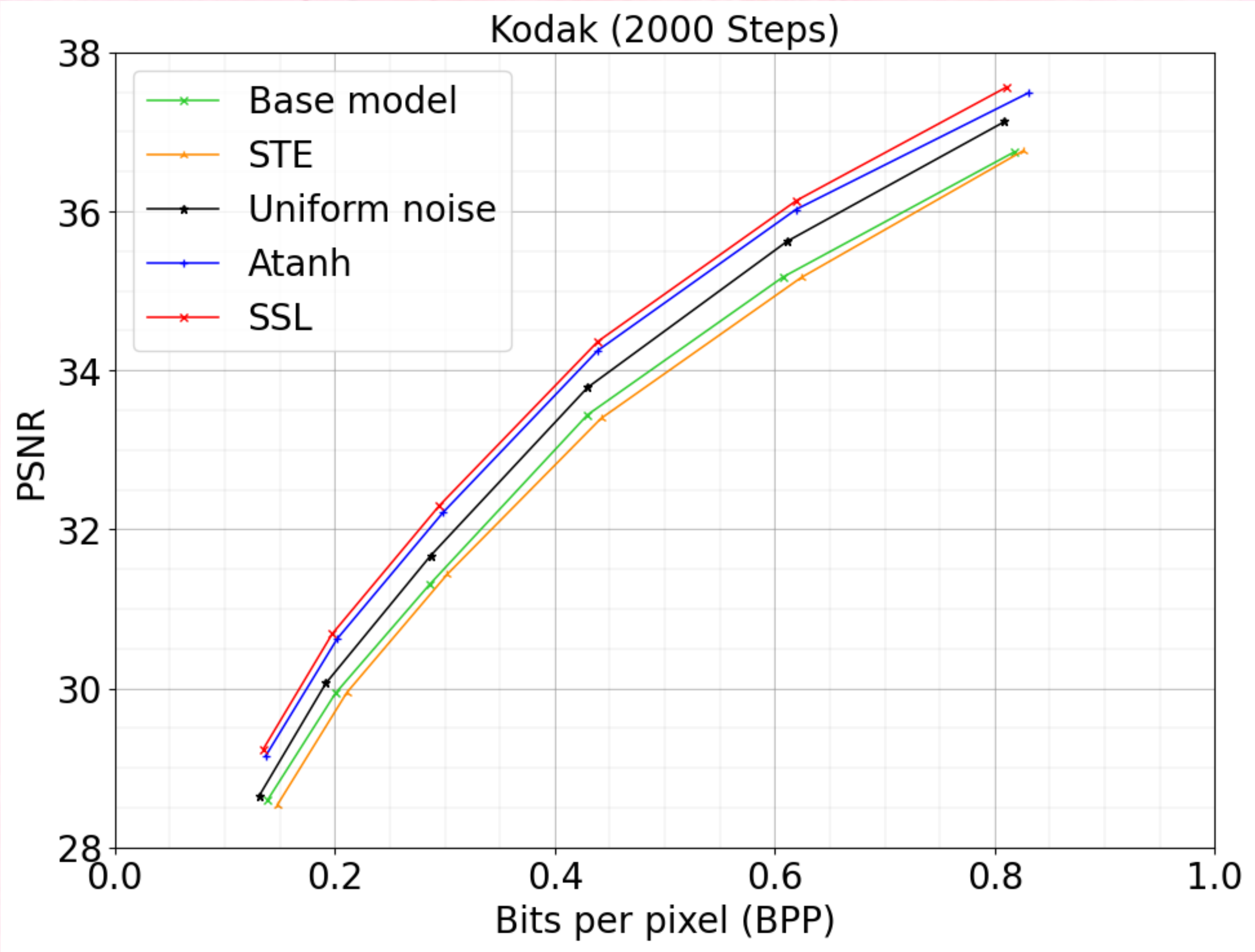

Figure 2) R-D performance for SSL on Kodak

Qualitative results

Original image

SSL (our method)

Temperature sensitivity

Table 1) True R-D loss for different $\tau_{\{\max\}}$ settings

Function \ τ_{max}	0.2	0.4	0.6	0.8	1.0
$\exp \operatorname{atanh}(v)$	0.6301	0.6273	0.6267	0.6260	0.6259
$1 - v$ (linear)	0.6291 ↓	0.6229 ↓	0.6225	0.6222	0.6220
$\cos^2(\frac{v\pi}{2})$	0.6307	0.6233	0.6194 ↓	0.6186	0.6187
$\sigma(-a\sigma^{-1}(v))$	0.6341	0.6233	0.6196	0.6181 ↓	**0.6175** ↓
$\exp \operatorname{atanh}(v)$	0.0010	0.0044	0.0073	0.0079	0.0084
$1 - v$ (linear)	0	0	0.0031	0.0041	0.0045

Conclusion

- Proposed SGA+ a more effective extension for refinement of latents and drop-in replacement for SGA
- SSL can interpolate between all proposed methods & outperforms all baselines in terms of R-D trade-off on Kodak, Tecnick and CLIC
- Exploration of SGA+ showed that it is more stable under varying conditions