Deep Stereo Image Compression with Decoder Side Information using Wyner Common Information

Ezgi Özyılkan

Supervisor: Prof. Deniz Gündüz

June 23, 2021

Our System Model

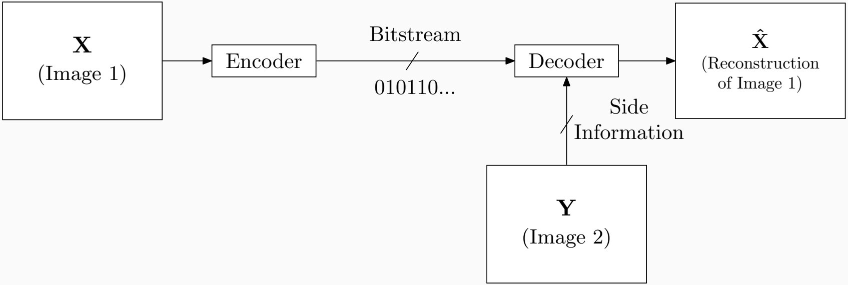


Figure 1: System model for lossy source compression with decoder-only side information. Also known as *Wyner-Ziv* model.

Non-linear Transform Coding

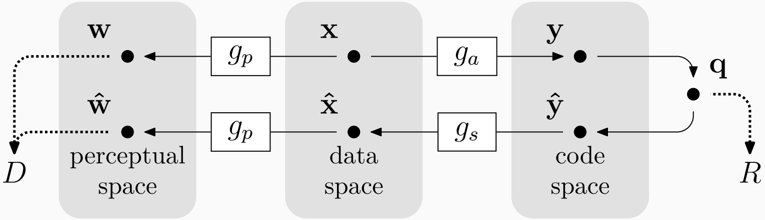


Figure 2: Nonlinear transform coding framework proposed in [Ballé et al., 2016]. The loss function we aim to minimize in this setting is denoted as $\mathrm{R} + \lambda D$. Figure provided is from [Ballé et al., 2017].

“Ballé2017” model

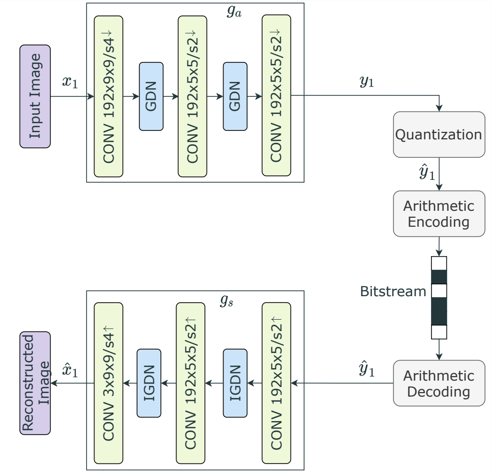


Figure 3: Proposed DNN-based image compression network architecture in [Ballé et al., 2017]. g_a and g_s blocks of layers in the figure refer to *analysis* and *synthesis* transforms from the data space, respectively.

“Ballé2018” model

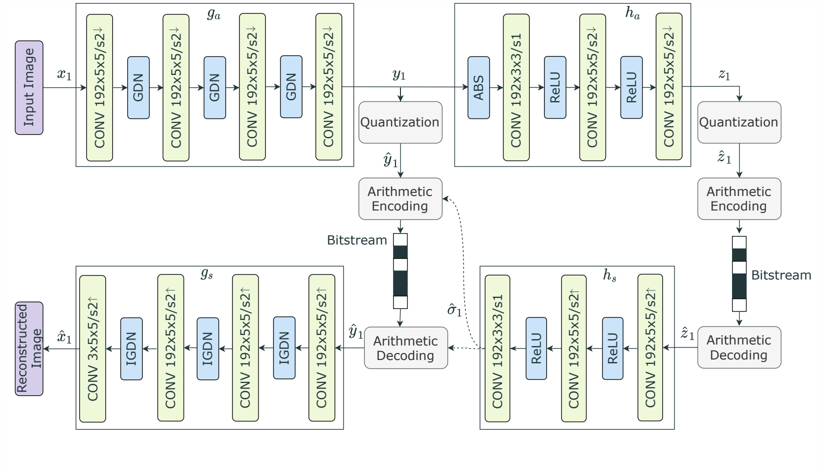


Figure 4: Proposed DNN-based image compression network architecture in [Ballé et al., 2018]. $\mathbf{h_a}$ and $\mathbf{h_s}$ blocks of layers in the figure refer to synthesis and analysis transforms related to the *hyperprior*, introduced as an extension to [Ballé et al., 2017].

Loss Function in [Ballé et al., 2017]

$$\begin{aligned}\mathbb{E}_{\mathrm{x}\sim p(\mathrm{x})} D_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\tilde{\mathrm{y}} \mid \mathrm{x}) \,||\, p(\tilde{\mathrm{y}} \mid \mathrm{x})\right] &= \mathbb{E}_{\mathrm{x}\sim p(\mathrm{x})} \mathbb{E}_{\tilde{\mathrm{y}}\sim q_{\boldsymbol{\phi}}(\tilde{\mathrm{y}}\mid\mathrm{x})}\left[\log q_{\boldsymbol{\phi}}(\tilde{\mathrm{y}} \mid \mathrm{x}) - \log p(\tilde{\mathrm{y}} \mid \mathrm{x})\right] \\ &= \mathbb{E}_{\mathrm{x}\sim p(\mathrm{x})} \mathbb{E}_{\tilde{\mathrm{y}}\sim q_{\boldsymbol{\phi}}(\tilde{\mathrm{y}}\mid\mathrm{x})}\Bigg[\log q_{\boldsymbol{\phi}}(\tilde{\mathrm{y}} \mid \mathrm{x}) \\ &\quad - \Bigg(\underbrace{\log p_{\boldsymbol{\theta}}(\mathrm{x} \mid \tilde{\mathrm{y}})}_{\text{distortion}} + \underbrace{\log p(\tilde{\mathrm{y}})}_{\text{rate}}\Bigg)\Bigg] + const.\end{aligned} \tag{1}$$

$$L(\mathbf{g}_a, \mathbf{g}_s) = R + \lambda D. \tag{2}$$

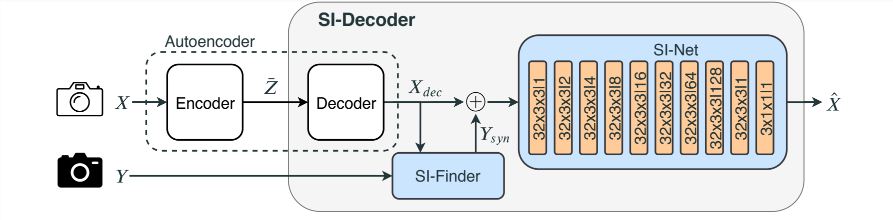


Figure 5: The DSC network architecture proposed in [Ayzik and Avidan, 2020].

$$L = (1-\alpha)\cdot d(X, X_{dec}) + \alpha \cdot d(X, \hat{X}) + \beta \cdot H(\bar{Z}) \quad (3)$$

Wyner Common Information

The concept of *Wyner common information* [Wyner, 1975] models the dependence between two random variables X and Y. It is defined as

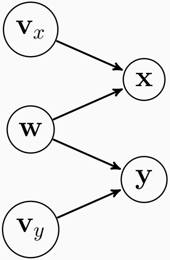
$$C(X;Y) = \inf I(X,Y;W), \tag{4}$$

where the infimum is taken over all the auxiliary random variables W such that $X - W - Y$ form a Markov chain.

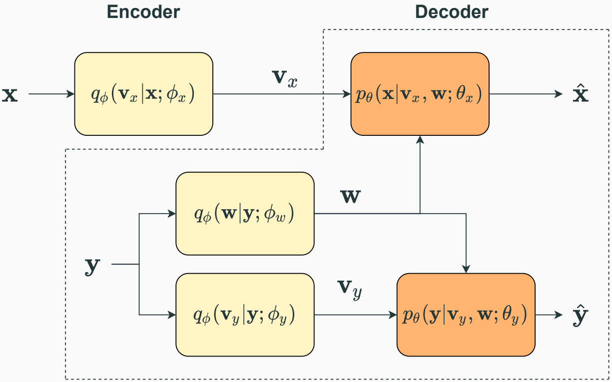
This implies:

- $p(w \mid x, y) = p(w \mid y)$
- $I(X, Y \mid W) = 0$

Distributed Source Coding (DSC) Architecture



Graphical model.



Distributed source coding architecture.

Figure 6: Proposed probability model and NN architecture.

Proposed Loss Function

$$
\begin{aligned}
&\mathbb{E}_{\mathsf{x},\mathsf{y}\sim p(\mathsf{x},\mathsf{y})} D_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\tilde{\mathsf{v}}_x, \mathsf{v}_y, \mathsf{w} \mid \mathsf{x}, \mathsf{y}) \mid\mid p(\tilde{\mathsf{v}}_x, \mathsf{v}_y, \mathsf{w} \mid \mathsf{x}, \mathsf{y})\right] \\
&= \mathbb{E}_{\mathsf{x},\mathsf{y}\sim p(\mathsf{x},\mathsf{y})} \mathbb{E}_{\tilde{\mathsf{v}}_x, \mathsf{v}_y, \mathsf{w} \sim q_{\boldsymbol{\phi}}} \Bigg(\Big(\log q_{\boldsymbol{\phi}}(\tilde{\mathsf{v}}_x \mid \mathsf{x}; \boldsymbol{\phi}_x) + \log q_{\boldsymbol{\phi}}(\mathsf{v}_y \mid \mathsf{y}; \boldsymbol{\phi}_y) + \log q_{\boldsymbol{\phi}}(\mathsf{w} \mid \mathsf{y}; \boldsymbol{\phi}_f) \Big) \\
&\qquad - \Big(\underbrace{\log p_{\boldsymbol{\theta}}(\mathsf{x} \mid \mathsf{w}, \tilde{\mathsf{v}}_x; \boldsymbol{\theta}_x)}_{D_x} + \underbrace{\log p_{\boldsymbol{\theta}}(\mathsf{y} \mid \mathsf{w}, \mathsf{v}_y; \boldsymbol{\theta}_y)}_{D_y} + \underbrace{\log p(\mathsf{w})}_{R_w} + \underbrace{\log p(\tilde{\mathsf{v}}_x)}_{R_x} + \underbrace{\log p(\mathsf{v}_y)}_{R_y} \Big) \Bigg)
\end{aligned} \tag{5}
$$

Derivation is included in the Appendix.

Overall Proposed Network Architecture

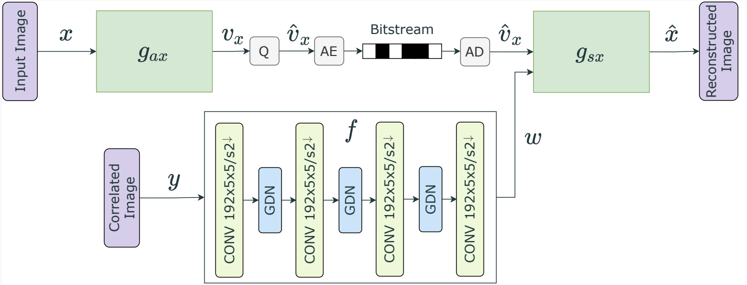


Figure 7: The proposed network architecture for distributed source coding. Block Q corresponds to a uniform quantizer, while blocks AE and AD correspond to arithmetic encoder and arithmetic decoder, respectively.

Similarly, we write:

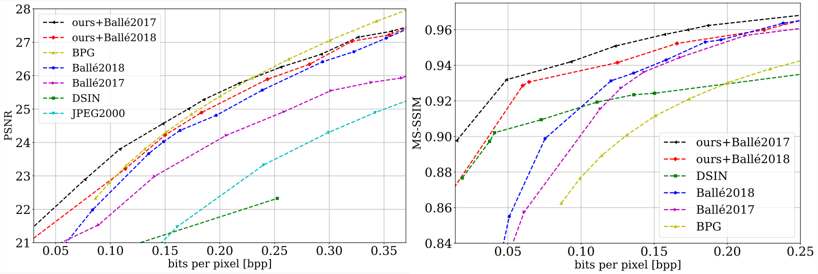
$$L(\mathbf{g}_{ax}, \mathbf{g}_{sx}, \mathbf{g}_{ay}, \mathbf{g}_{sy}, \mathbf{f}) = (R_x + \lambda D_x) + \alpha\,(R_y + \lambda D_y) + \beta R_w, \qquad (6)$$

Experimental Setup

KITTI Stereo

Figure 8: Example image pair from KITTI Stereo.

Results



(a) (b)

Figure 9: Comparison of different models in terms of MSE and MS-SSIM metrics.

Visual Comparisons i

Figure 10: Visual comparison of different models trained for the MS-SSIM metric. "Ours" in the figures above refers to the "Ours + Ballé2017" model.

Visual Comparisons ii

(a) DSIN, bpp=0.0449

(b) Ours, bpp=0.0431

Figure 11: Reconstruction comparison between DSIN (top) and ours (bottom) when evaluated on full-sized images from KITTI Stereo.

Conclusion and Final Words

- Addressing the gap in ML literature by offering information-theoretic foundation for the DSC problem
- Future directions include:
  - Two distributed encoders
  - Enhancing image quality at higher bpps
- Code publicly available at: *https://github.com/ipc-lab/DWSIC*

Acknowledgements[1]

- Dr. Nitish Mital, Imperial College
- Ali Garjani, Sharif University of Technology
- Prof. Deniz Gündüz, Imperial College

[1] N. Mital*, E. Ozyilkan*, A. Garjani*, and D. Gunduz. Deep Stereo Image Compression with Decoder Side Information using Wyner Common Information. arXiv:2106.11723.

Questions?

Email: `ezgi.ozyilkan17@imperial.ac.uk`

Appendix i

$$\text{p(x,y,w,v}_x, \mathbf{v}_y) = p(\mathbf{w})p(\mathbf{v}_x)p(\mathbf{v}_y)p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{w}, \mathbf{v}_x; \boldsymbol{\theta}_x)p_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{w}, \mathbf{v}_y; \boldsymbol{\theta}_y),$$

Joint distribution of the random variables.

$$q_{\boldsymbol{\phi}}(\mathbf{w}, \mathbf{v}_x, \mathbf{v}_y \mid \mathbf{x}, \mathbf{y}) = q_{\boldsymbol{\phi}}(\mathbf{v}_x \mid \mathbf{x}; \phi_x)q_{\boldsymbol{\phi}}(\mathbf{w} \mid \mathbf{y}; \phi_w)q_{\boldsymbol{\phi}}(\mathbf{v}_y \mid \mathbf{y}; \phi_y), \quad (7)$$

Factored variational approximation of the posterior distribution[2].

[2] Eq. (16) and (7) are similar to the ones proposed in [Wang et al., 2017].

Derivation for Eq. (5).

$$
\begin{aligned}
&\mathbb{E}_{x,y\sim p(x,y)} D_{\mathrm{KL}}\left[q_{\phi}(\tilde{v}_x, v_y, w \mid x, y) \,||\, p(\tilde{v}_x, v_y, w \mid x, y)\right] \\
&= \mathbb{E}_{x,y\sim p(x,y)} \mathbb{E}_{\tilde{v}_x, v_y, w \sim q_{\phi}}\left[\log\left(\frac{q_{\phi}(\tilde{v}_x, v_y, w \mid x, y)}{p(\tilde{v}_x, v_y, w \mid x, y)}\right)\right] && (8)\\
&= \mathbb{E}_{x,y\sim p(x,y)} \mathbb{E}_{\tilde{v}_x, v_y, w \sim q_{\phi}}\left[\log\left(\frac{q_{\phi}(\tilde{v}_x, v_y, w \mid x, y) p(x,y)}{p(\tilde{v}_x, v_y, w, x, y)}\right)\right] && (9)\\
&= \mathbb{E}_{x,y\sim p(x,y)} \mathbb{E}_{\tilde{v}_x, v_y, w \sim q_{\phi}}\left[\log\left(\frac{q_{\phi}(\tilde{v}_x \mid x; \phi_x) q_{\phi}(v_y \mid y; \phi_y) q_{\phi}(w \mid y; \phi_f) p(x,y)}{p(x \mid \tilde{v}_x, w; \theta_x) p(y \mid v_y, w; \theta_y) p(w) p(\tilde{v}_x) p(v_y)}\right)\right] && (10)\\
&= \mathbb{E}_{x,y\sim p(x,y)} \mathbb{E}_{\tilde{v}_x, v_y, w \sim q_{\phi}}\Bigg(\Big(\log q_{\phi}(\tilde{v}_x \mid x; \phi_x) + \log q_{\phi}(v_y \mid y; \phi_y) + \log q_{\phi}(w \mid y; \phi_f)\Big) \\
&\qquad - \Big(\underbrace{\log p_{\theta}(x \mid w, \tilde{v}_x; \theta_x)}_{D_x} + \underbrace{\log p_{\theta}(y \mid w, v_y; \theta_y)}_{D_y} + \underbrace{\log p(w)}_{R_w} + \underbrace{\log p(\tilde{v}_x)}_{R_x} + \underbrace{\log p(v_y)}_{R_y}\Big)\Bigg) \\
&\qquad + \mathit{const}, && (11)
\end{aligned}
$$

where Eq. (10) follows from Eq. (16) and (7).

References

- Ayzik, S., & Avidan, S. (2020). Deep image compression using decoder side information. *European Conf. on Computer Vision - ECCV*, 699–714.
- Ballé, J., Laparra, V., & Simoncelli, E. P. (2017). End-to-end optimized image compression. *Int'l Conf on Learning Representations (ICLR)*.
- Ballé, J., Minnen, D., Singh, S., Hwang, S. J., & Johnston, N. (2018). Variational image compression with a scale hyperprior. *International Conference on Learning Representations.*
- Ballé, J., Laparra, V., & Simoncelli, E. P. (2016). End-to-end optimization of nonlinear transform codes for perceptual quality.

Wang, W., Yan, X., Lee, H., & Livescu, K. (2017). Deep variational canonical correlation analysis.

Wyner, A. (1975). The common information of two dependent random variables. *IEEE Transactions on Information Theory*, *21*(2), 163–179.