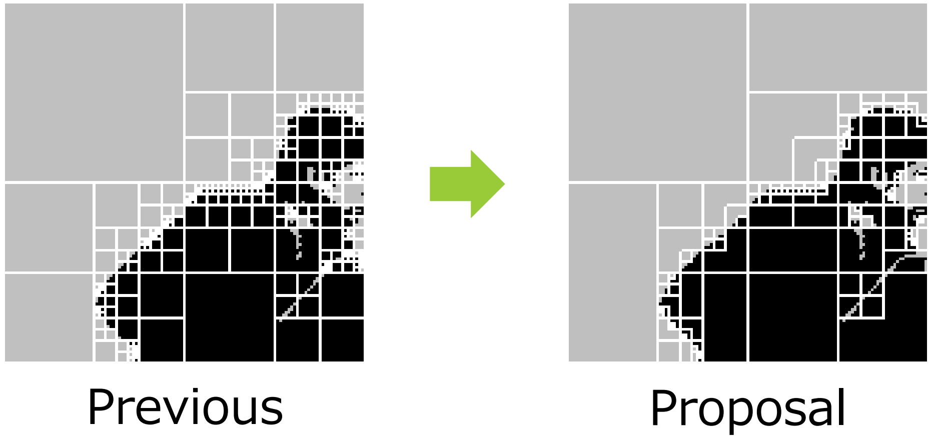
Stochastic Model of Block Segmentation Based on Improper Quadtree and Optimal Code under the Bayes Criterion

Yuta Nakahara[1] and Toshiyasu Matsushima[2]
1) Center for Data Science, Waseda University
2) Dept. of Pure and Applied Math. Waseda University

DCC 2022

Summary

- Our approach to lossless image compression
  - Assuming stochastic generative models directly on pixel values
  - Achieving the theoretical limit of the assumed models
- Contribution
  - Proposal of a stochastic model based on improper quadtrees
- Results
  - Efficient representation with fewer regions (= fewer parameters)
  - Improvement of code lengths in lossless compression



Outline

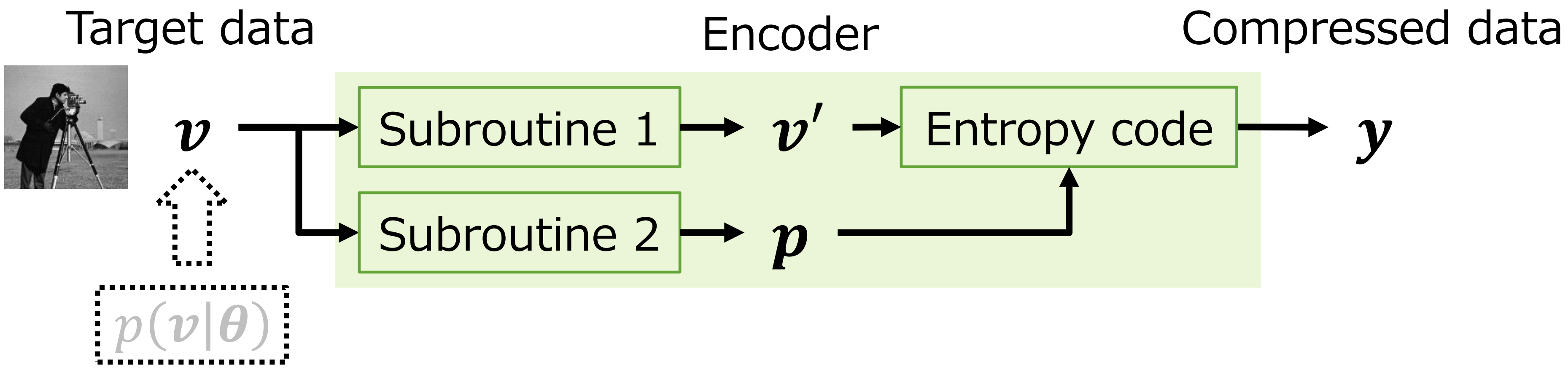
- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

Outline

- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

Lossless compression as an image processing

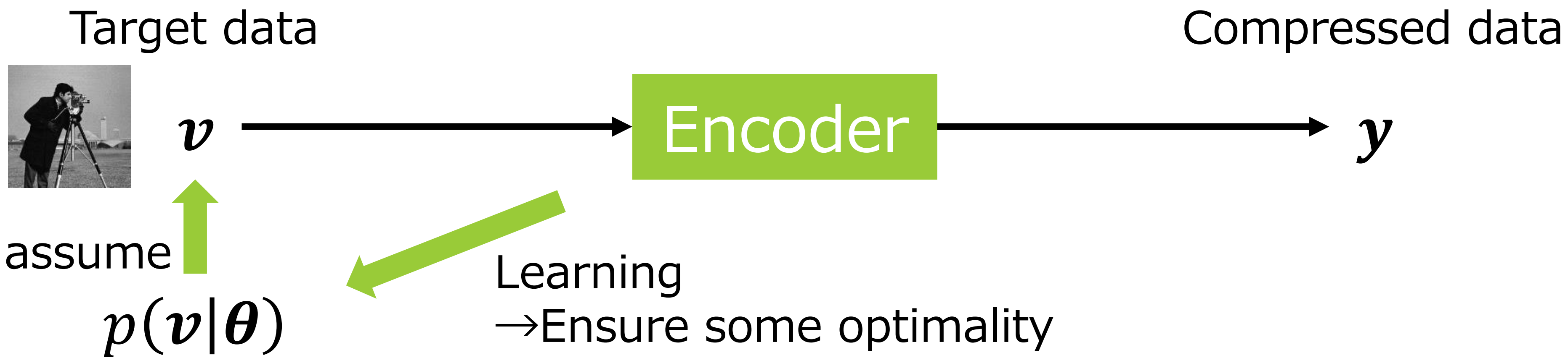
- Coding procedure



- Assumption of $p(\boldsymbol{v}|\boldsymbol{\theta})$ is implicit.
- $\boldsymbol{p}$ does not directly govern the generation of $\boldsymbol{v}$.
(Just a vector s.t. $\sum_i p_i = 1$)
- Difficulty in discussing the optimality of the subroutines to $p(\boldsymbol{v}|\boldsymbol{\theta})$.

Lossless compression on stochastic generative models

- Problem setup



- $\boldsymbol{\theta}$ governs the probabilistic generation of $\boldsymbol{v}$.
  ⇒$\boldsymbol{\theta}$ can be statistically learned from $\boldsymbol{v}$.
  (Implicit correspondence with Subroutine 2)
- Bayes codes[Matsushima et al, 1991]:
  the codes with Bayesian learning of $\boldsymbol{\theta}$

Previous studies – redefinition of stochastic generative models

- Some subroutines play a role like parameter learning for an implicitly assumed $p(\boldsymbol{x}|\boldsymbol{\theta})$
- We explicitly redefine the implicit $p(\boldsymbol{x}|\boldsymbol{\theta})$ and derive the optimal coding algorithm for it.

Subroutine	Stochastic generative model
Linear prediction (e.g., [Kuroki et al, 1992], [Wu et al, 1998])	Linear autoregressive model [Nakahara and Matsushima, 2020]
Quadtree block segmentation [Matsuda et al, 2005]	Stochastic quadtree model [Nakahara and Matsushima, 2020]
Edge prediction [Meyer, 1997]	2D-HMM [Nakahara and Matsushima, 2020]
Predictor weighting and switching (e.g., [Weinberger et al, 2000], [Martchenko and Deng, 2013])	Mixture of priors [Nakahara and Matsushima, 2021]

➡ Extend in this study (Stochastic quadtree model)

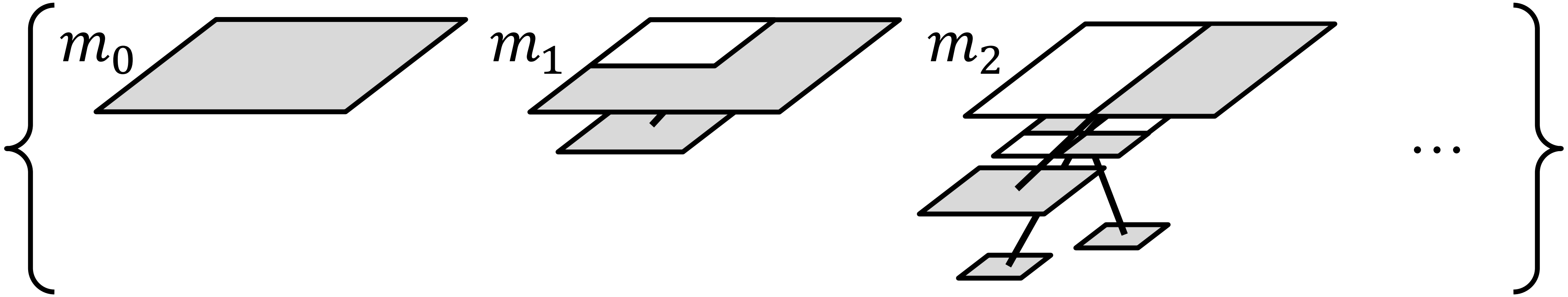
Outline

- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

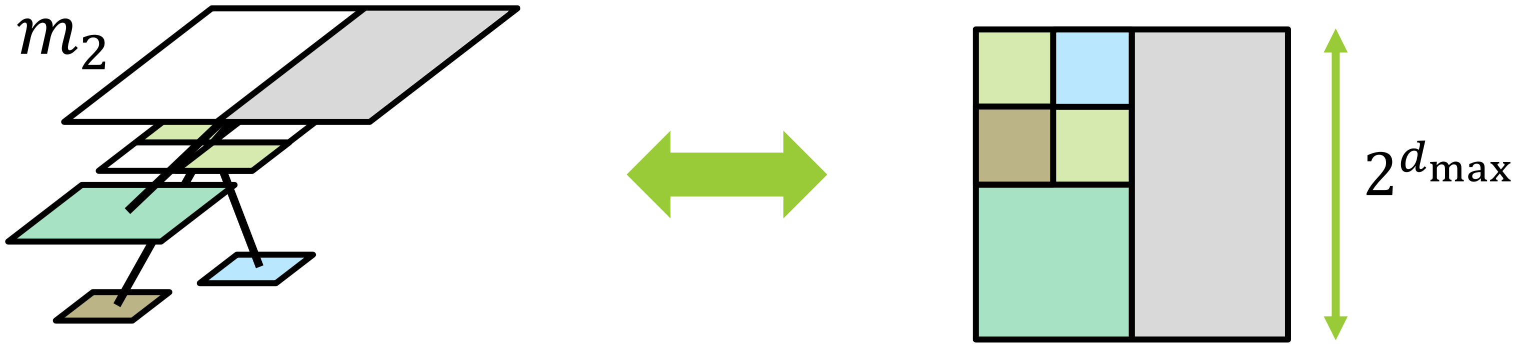
Proposed stochastic generative model

- Let both width and height are $2^{d_{\max}}$.
- The set of the improper quadtrees whose depth $\leq d_{\max}$.

$\mathcal{M} =$

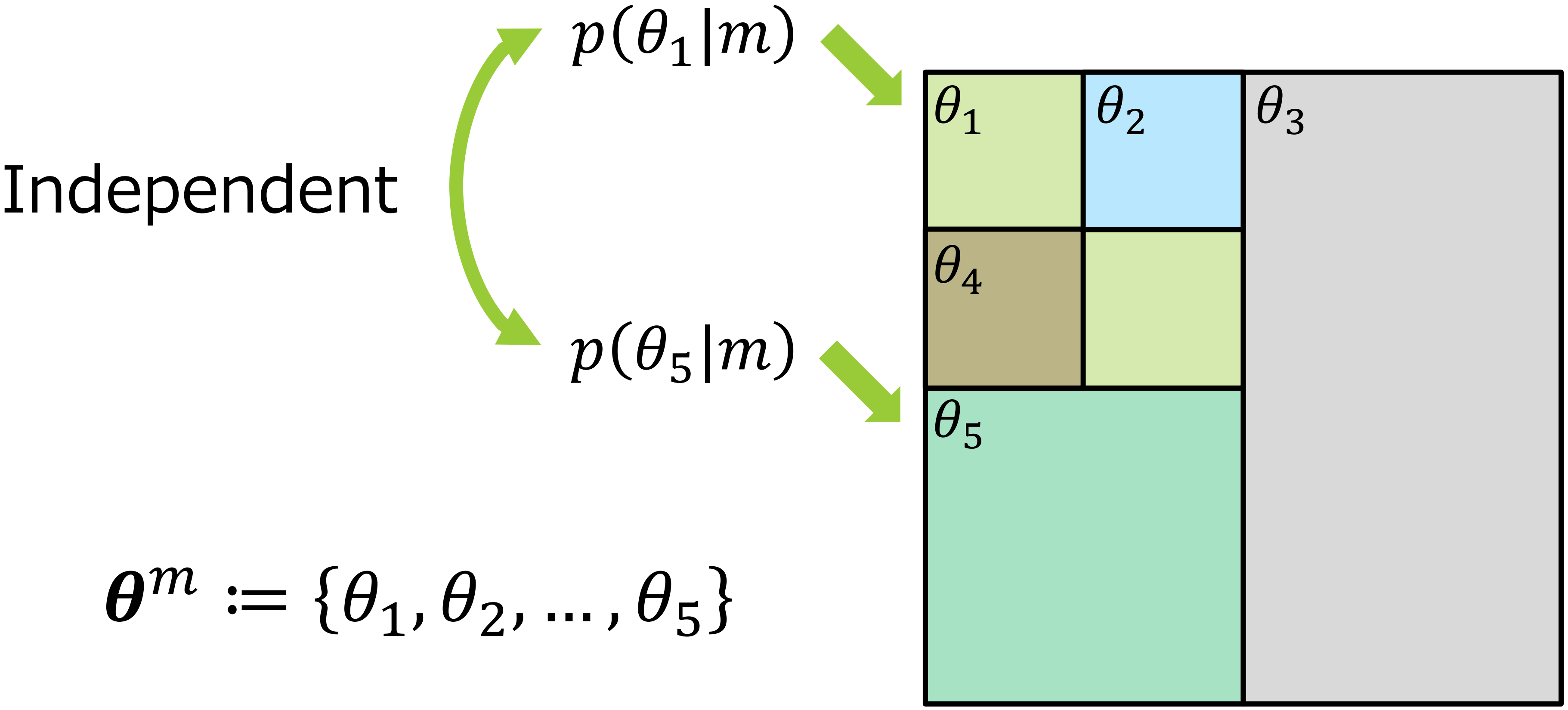


- One of them is chosen with probability $p(m)$.



Proposed stochastic generative model

- Parameter θ_s is independently assigned to each block s with probability $p(\theta_s|m)$.

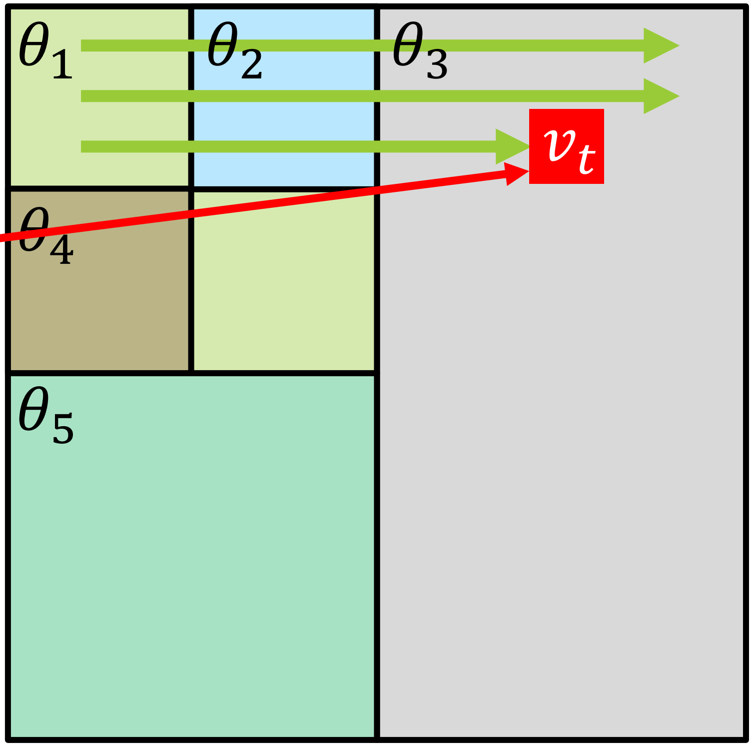


$$\boldsymbol{\theta}^m := \{\theta_1, \theta_2, \dots, \theta_5\}$$

Proposed stochastic generative model

- Pixel value v_t at block s is generated in order of the raster scan with probability $p(v_t|v^{t-1},\theta_s,m)$.

$$p(v_t|v^{t-1},\theta_3,m)$$



v_t depends only on
- The past sequence v^{t-1}
- The parameter θ_s of the block s which contains v_t

Outline

- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

The Bayes codes [Matsushima et al, 1991]

- We cannot use $p(v_t|v^{t-1},\boldsymbol{\theta}^m,m)$ because true m and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner. ⇒ Bayes codes

Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1}) = \sum_{m\in\mathcal{M}} p(m|v^{t-1}) \int p(v_t|v^{t-1},\boldsymbol{\theta}^m,m)p(\boldsymbol{\theta}^m|v^{t-1},m)\,\mathrm{d}\boldsymbol{\theta}^m$$

Bayes codes

- Its expected code length converges to the entropy for sufficiently large data length[Clarke and Barron, 1990].
- Its convergence speed achieves the theoretical limit [Clarke and Barron, 1990].
- Efficient text coding algorithms have been constructed based on it (e.g., [Matsushima and Hirasawa, 2009]).

The Bayes codes [Matsushima et al, 1991]

- We cannot use $p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m)$ because true m and $\boldsymbol{\theta}^m$ are unknown.
- We estimate it by $\hat{p}_c(v_t|v^{t-1})$ in Bayesian manner. ⇒ Bayes codes

Optimal coding probability $p_c^*(v_t|v^{t-1})$ for our model

$$p_c^*(v_t|v^{t-1}) = \underline{\sum_{m\in\mathcal{M}}} \underline{p(m|v^{t-1})} \underline{\int p(v_t|v^{t-1}, \boldsymbol{\theta}^m, m) p(\boldsymbol{\theta}^m|v^{t-1}, m) \, \mathrm{d}\boldsymbol{\theta}^m}$$

Three computationally hard parts

- 1. The summation w.r.t. m ← A recursive structure of quadtree
- 2. The posterior $p(m|v^{t-1})$ ← Special prior (Detailed in the paper)
- 3. The integral w.r.t. $\boldsymbol{\theta}^m$ ← Conjugate prior

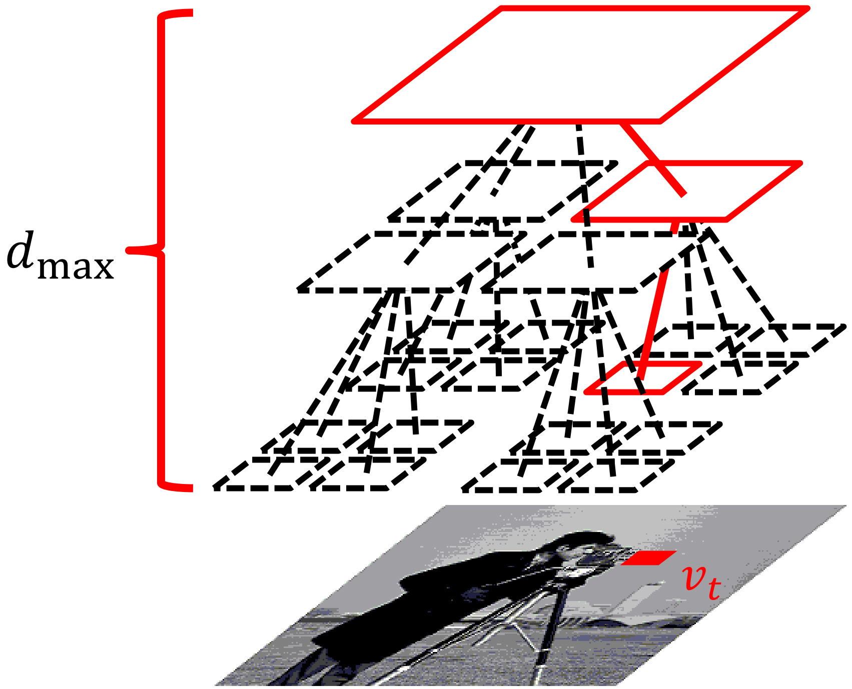
Proposed Algorithm

- Our algorithm reduces the complexity from $O(|\mathcal{M}|)$ to $O(2^4 d_{\max})$ **without any approximation**.

Doubly exponential w.r.t. $d_{\max}$
E.g., $d_{\max} = 2 \Rightarrow |\mathcal{M}| = 83{,}521$

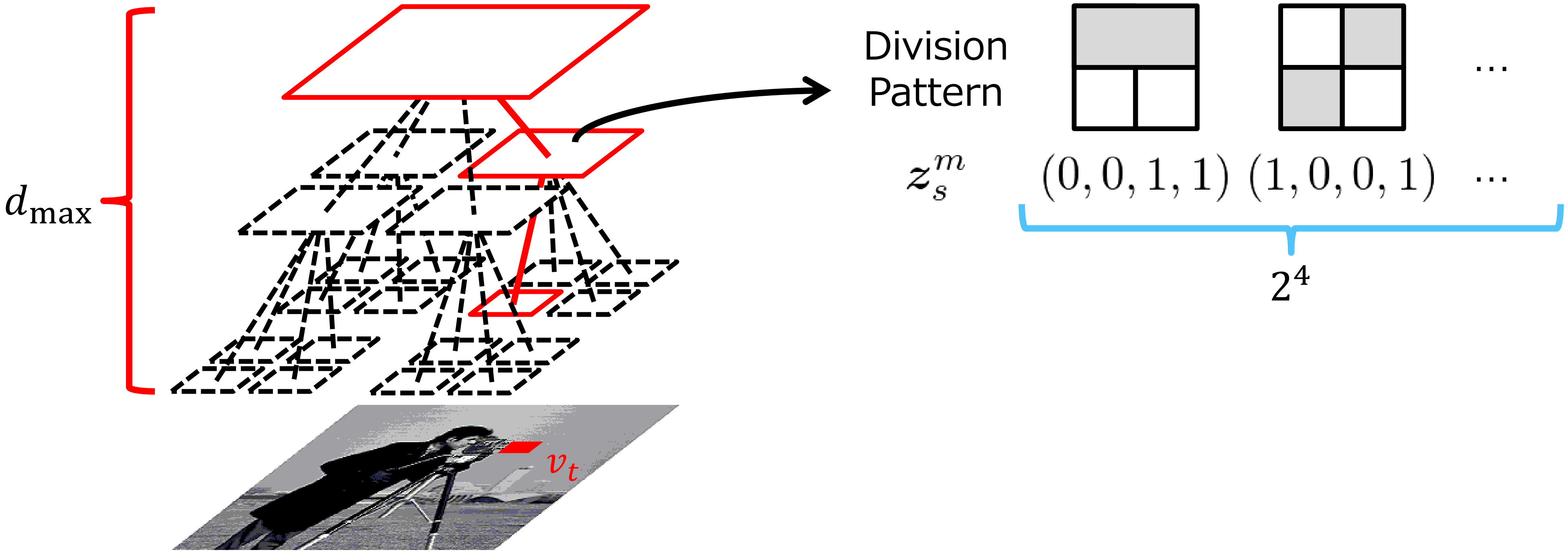
Proposed Algorithm

- Our algorithm reduces the complexity from $O(|\mathcal{M}|)$ to $O(2^4 d_{\max})$ **without any approximation**.
- Factorization based on the independency among the regions



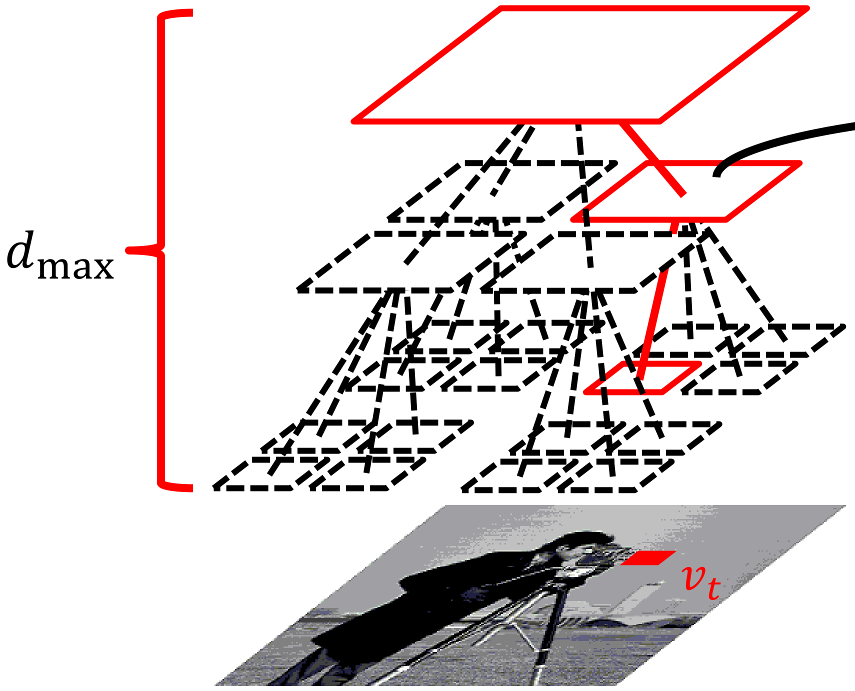
Proposed Algorithm

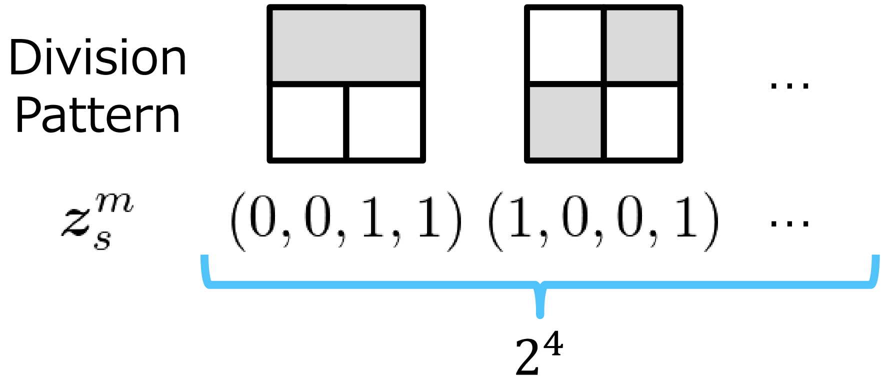
- Our algorithm reduces the complexity from $O(|\mathcal{M}|)$ to $O(2^4 d_{\max})$ **without any approximation**.
- Factorization based on the independency among the regions



Proposed Algorithm

- Our algorithm reduces the complexity from $O(|\mathcal{M}|)$ to $O(2^4 d_{\max})$ **without any approximation**.
- Factorization based on the independency among the regions





$$q(v_t|v^{t-1},s) = \sum_{\mathbf{z}_s : z_{s s_{\mathrm{ch}}}=0} \eta_s(\mathbf{z}_s|v^{t-1}) \tilde{q}(v_t|v^{t-1},s,\mathbf{z}_s) + \left(\sum_{\mathbf{z}_s : z_{s s_{\mathrm{ch}}}=1} \eta_s(\mathbf{z}_s|v^{t-1}) \right) q(v_t|v^{t-1},s_{\mathrm{ch}})$$

Outline

- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

Experiment 1

- Purpose: confirmation of the Bayes optimality
- Setting:
  - We generated 1000 binary images (64x64) as follows:
    1. Generate m according to $p(m)$ (detailed in the paper).
    2. Generate θ_s according to $\mathrm{Beta}(\theta_s|\alpha,\beta)$ for each block s.
    3. Generate v_t according to $\mathrm{Bern}(v_t|\,\theta_s)$ for each block s.
  - We compressed them by
    - The proposed method with improper quadtrees
    - The method with proper quadtrees
    - JBIG (implementation by [Kuhn, 1995])

Experiment 1

Result:

- Examples of the generated images

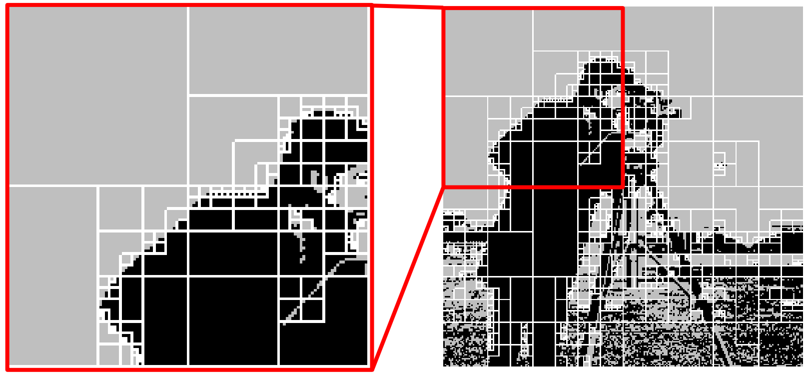
- Average coding rates (bit/pel)

Improper QT (proposal)	Proper QT	JBIG
0.619	0.624	1.811

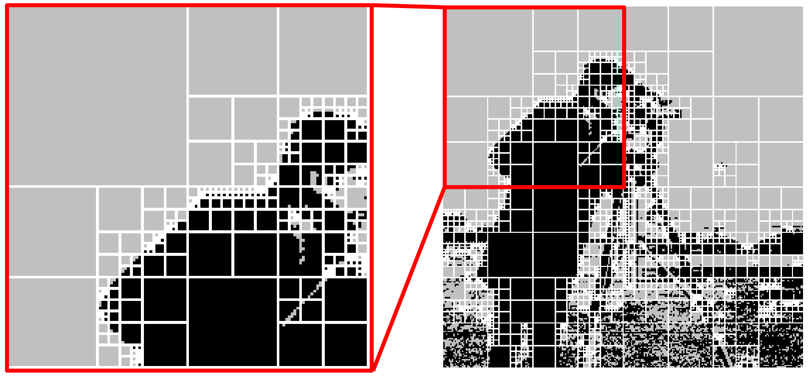
Experiment 2

- Purpose: Confirmation of the suitability to real images
- Result:
  - MAP estimated model $m^{\mathrm{MAP}} = \mathrm{argmax}_{m \in \mathcal{M}} p(m|\boldsymbol{v})$.

Improper QT (proposal)



Proper QT



  - Average coding rates (bit/pel)

Improper QT (proposal)	Proper QT	JBIG
0.318	0.323	0.348

Outline

- Background
  - Lossless compression as an image processing
  - Lossless compression on stochastic generative models
- Previous studies – redefinition of stochastic generative models
- Proposed stochastic generative models
- Bayes codes
- Proposed algorithm
- Experiments
- Conclusion

Conclusion

- We assumed the stochastic generative model directly on pixel values and achieved the theoretical limit of the assumed model.
- The proposed stochastic model was based on improper quadtrees.
- We obtained the efficient representation with fewer regions and improved the average code length in lossless image compression.

References

- T. Matsushima, H. Inazumi, and S. Hirasawa, "A class of distortionless codes designed by Bayes decision theory," IEEE Transactions on Information Theory, vol. 37, no. 5, pp. 1288-1293, Sep. 1991.
- N. Kuroki, T. Nomura, M. Tomita, and K. Hirano, "Lossless image compression by two-dimensional linear prediction with variable coefficients," IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, vol. 75, no. 7, pp. 882-889, 1992.
- X. Wu, E. Barthel, and W. Zhang, "Piecewise 2d autoregression for predictive image coding," in Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269), 1998, pp. 901-904.
- I. Matsuda, N. Ozaki, Y. Umezu, and S. Itoh, "Lossless coding using variable block-size adaptive prediction optimized for each image," in 2005 13th European Signal Processing Conference, 2005, pp. 1-4.
- B. Meyer and P. Tischer, "TMW - a new method for lossless image compression," in Proc. of 1997 Picture Coding Symposium (PCS'97), 1997, pp. 533-538.

References

- M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEGLS," IEEE Transactions on Image Processing, vol. 9, no. 8, pp. 1309-1324, Aug 2000.
- A. Martchenko and G. Deng, "Bayesian predictor combination for lossless image compression," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5263-5270, 2013.
- Y. Nakahara and T. Matsushima, "Autoregressive image generative models with normal and t-distributed noise and the Bayes codes for them," in 2020 International Symposium on Information Theory and Its Applications (ISITA), 2020, pp. 81-85.
- Y. Nakahara and T. Matsushima, "A stochastic model for block segmentation of images based on the quadtree and the Bayes code for it," Entropy, vol. 23, no. 8, 2021.
- Y. Nakahara and T. Matsushima, "Bayes code for two-dimensional auto-regressive hidden Markov model and its application to lossless image compression," in International Workshop on Advanced Imaging Technology (IWAIT) 2020, vol. 11515. SPIE, 2020, pp. 330-335. [Online]. Available: https://doi.org/10.1117/12.2566943

References

- Y. Nakahara and T. Matsushima, "Hyperparameter Learning of Stochastic Image Generative Models with Bayesian Hierarchical Modeling and Its Effect on Lossless Image Coding," 2021 IEEE Information Theory Workshop (ITW), 2021, pp. 1-6, doi: 10.1109/ITW48936.2021.9611418.
- B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," IEEE Transactions on Information Theory, vol. 36, no. 3, pp. 453-471, May 1990.
- T. Matsushima and S. Hirasawa, "Reducing the space complexity of a Bayes coding algorithm using an expanded context tree," in 2009 IEEE International Symposium on Information Theory, June 2009, pp. 719-723.
- M. Kuhn, "JBIG-KIT," https://www.cl.cam.ac.uk/~mgk25/jbigkit/, Accessed: 2022-3-5.