8-2020

Machine Learning Enhanced Free-Space and Underwater OAM Optical Communications

Patrick L. Neary
Utah State University

Recommended Citation

Neary, Patrick L., "Machine Learning Enhanced Free-Space and Underwater OAM Optical Communications" (2020). *All Graduate Theses and Dissertations*. 7892.
https://digitalcommons.usu.edu/etd/7892

MACHINE LEARNING ENHANCED FREE-SPACE AND UNDERWATER OAM OPTICAL COMMUNICATIONS

by

Patrick L. Neary

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

Nicholas Flann, Ph.D. Major Professor	Vicki Allan, Ph.D. Committee Member
Vladimir Kulyukin, Ph.D. Committee Member	Stephen Clyde, Ph.D. Committee Member
Tyler Brough, Ph.D. Committee Member	Richard S. Inouye, Ph.D. Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

ABSTRACT

Machine Learning Enhanced Free-Space and Underwater OAM Optical Communications

by

Patrick L. Neary, Doctor of Philosophy
Utah State University, 2020

Major Professor: Nicholas Flann, Ph.D.
Department: Computer Science

Machine learning has wound its way into a broad range of areas. Among those applications is the exploratory field of free-space and underwater optical communications using orbital angular momentum (OAM) of light. With proper hardware, a laser beam can go from a Gaussian shaped distribution to a doughnut shaped pattern, with a configurable radius. Multiple OAM patterns, or modes, can be multiplexed to create unique interference patterns. OAM patterns can be used to encode bits for communicating information.

Traditional approaches in optical demultiplexing require hardware that is sensitive to alignment and other environmental factors. Using machine learning to identify patterns allows the de-multiplexing of OAM patterns without complicated or expensive hardware. In lab environments signal integrity can be ensured, however, operational integrity in non-ideal environments can cause serious signal degradation.

This work explores ways to improve pattern recognition or classification in both underwater and free-space environments. Specifically, principles of physics inspired training are applied to convolutional neural networks to make the trained networks more robust to signal attenuation. Continuous, linear optimal transport principles are applied to improve the classification of OAM patterns. Optimal transport is used to create attenuation models that, when inserted in the training pipeline, improve neural network robustness to signal attenuation. Finally, state-of-the-art deep convolutional neural networks are explored to see which provide the most robust performance in free-space and underwater communications. These environments are explored when the signal is degraded by turbulence and attenuation.

(171 pages)

PUBLIC ABSTRACT

Machine Learning Enhanced Free-Space and Underwater OAM Optical Communications

Patrick L. Neary

Communications, bandwidth, security, and hardware simplicity are principles of interest to society at large. Recent advances in optics and in understanding properties of light, such as orbital angular momentum (OAM), have provided new potential mediums for communication.

Machine learning has wound its way into a broad range of fascinating areas. An emerging field of research is the use of a unique property of lasers called orbital angular momentum (OAM). With the proper hardware, a laser can go from a Gaussian shaped distribution to a doughnut shaped pattern, where the radius can be changed. Multiple OAM patterns, or modes, can be combined to create unique patterns. This research explores the use of machine learning to de-multiplex OAM patterns. The OAM patterns can be used to encode bits for communication.

This work explores ways to improve pattern recognition or classification in both underwater and free-space environments. Specifically, various approaches are applied to train convolutional neural networks to make them more robust to signal degradation through turbulence and attenuation. A new image transform is used to improve OAM pattern classification. Finally, some of the state of the art deep convolutional neural networks are explored to see which provide the most robust performance in free-space and underwater communications. A variety of methods are shown to improve the state of the art in pattern classification in OAM communications.

For my wife, Cynthia, who supported me through this long process and took the brunt of running our household while I was locked away in my office.

ACKNOWLEDGMENTS

To the Space Dynamics Laboratory, who so graciously provided the fellowship for making this all possible. To the Naval Research Laboratory, through their 6.1 and 6.2 base programs for OAM underwater communications, that provided resources, facilities, and data. To Dr. Vicki Allan, who was so supportive and encouraged me to find something I was truly passionate about. To Dr. Nicholas Flann who helped crystallize my focus. To my other committee members who taught classes that broadened my mind and deepened my knowledge. And especially to Cynthia and my boys who patiently supported me during this process.

Patrick L. Neary

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

ACRONYMS

AD	automatic differentiation
AL	attenuation length
ANN	artificial neural network
CNN	convolutional neural network
CDT	cumulative distribution transform
ConvNet	convolutional neural network
DFT	discriminative fine tuning
EM	electromagnetic
FCL	fully connected layer
FSO	free-space optical (communications)
GD	gradient descent
FSO	free space optical
LDA	linear discriminant analysis
LG	Laguerre-Gaussian
ML	machine learning
MSE	mean squared errors
NN	neural network
NS	nearest subspace
OAM	orbital angular momentum
OT	optimal transport
PCA	principal component analysis
PDE	partial differential equation
R-CDT	radon-cumulative distribution transform
SGD	stochastic gradient descent
SLM	spatial light modulator
SMART	SNR machine learning attenuation model
SNR	signal to noise ratio
SSE	sum of squared errors
UWOC	underwater optical communications

CHAPTER 1

INTRODUCTION

Communication mediums provide an important backbone to our society. Social media, entertainment, information distribution, information storage and retrieval, military communications, GPS, etc. all require a medium for transfer. Many communications take place over wire, fiber, and RF links.

A current estimate is that only 58% of the world's population is connected to the Internet (1). As Internet access to remote areas improves and usage continues to increase, demand for bandwidth will grow. A variety of difficulties exist for providing access to remote areas, including labor costs for laying cables, infrastructure hardware costs, and addressing physical obstacles (in water or land). Communication security is another are of consideration, such as eavesdropping of signals (2).

One approach to addressing these issues is the use of free-space optical (FSO) or underwater optical communications (UWOC). FSO and UWOC are different from RF communications in that they are based on optical, direct line of sight, transmission between a transmitter and receiver. Optical wavelengths can include ultraviolet, visible, and infrared light.

Free-space and underwater optical communications have a potential benefit of bypassing costly cable installations through geographically difficult regions. Another benefit is the fact that it is difficult to intercept the communication stream without degrading the signal and alerting the system to tampering attempts. However, security is an ongoing area of interest and study with optical communications (2). Information capacity needs will only continue to grow as more of the world becomes connected and more data is generated. Optical communications provide a potential solution to bandwidth bottlenecks (3).

The research in this dissertation is developed within the domain of optical communications. While the computer science (CS) and machine learning (ML) principles may be familiar to those in the CS field, the physics and optics details may be new. In light of this, several appendices have been added to provide background on a number of relevant topics. These may be read according to the interest and curiosity of the reader. Appendix A provides high level background on OAM, details on how the data was generated, and how it was divided for training and testing. Appendix B provides some of the math behind the wave equation, which is foundational to the starting point in many OAM related papers. Appendix C provides background on optimal transport and the radon-cumulative distribution transform (R-CDT). Appendix D contains a discussion of many of the state of the art CNNs and what makes them unique. Finally, Appendix E contains a primer on neural networks. Moving forward, familiarity with this background material is assumed.

1.1 Problem Overview

Orbital Angular Momentum (OAM) is a relatively new discovery as a property of coherent light. In 1992, Allen et al. discovered that Laguerre-Gaussian beams could transition from a forward propagating wavefront to a helical shaped path(4). Different modes can be applied to a beam, which cause the radius associated with helical path to increase or decrease (4).

While the fact that light can travel in a helical path is, in and of itself fascinating, there are some important properties that accompany this phenomenon. It turns out that OAM modes are completely orthogonal to each other (4), so different OAM modes can be multiplexed for transmission and be completely de-multiplexed upon receipt (5; 6). Initial tests using these ideas, in controlled environments, have resulted in terabit data rates under ideal conditions (5). OAM as a means for communication has created a significant stir in the research community (7; 8).

For OAM pattern classification in communications, there is a confluence of a number of interesting areas. These topics include OAM communications, convolutional neural network architectures, modeling, physics-inspired training, and automatic differentiation. In this section, literature in these topics will be reviewed with a special emphasis on the subset of those areas that relate to the subsequent research. Once the review of relevant literature is complete, specific research objectives and contributions will be outlined.

OAM, via electromagnetic waves, has shown great potential in communications research (7; 8). OAM communications can take place in either free-space (9) or underwater (10) environments and communication bandwidth can be increased by multiplexing modes together. Several significant degrading factors in OAM communications exist, these include turbulence (11) and signal attenuation (12).

The benchtop OAM communication setup used for transmitting and receiving OAM beams in this work is shown in Figure 1.1. A laser generates a Gaussian shaped beam, which is reflected off a mirror and passed through a series of beamsplitters. The light from each beamsplitter is passed through phase plates where a unique OAM mode is imparted to each beam. The phase plates can be enabled or disabled to allow for different combinations of the patterns. After passing through the phase plates, the beams are recombined using beamsplitters and the multiplexed beams are then passed through a tank (for underwater communications). The tank has pumps that keep attenuation particles in suspension. As the beam exits the tank, mirrors are used to direct the beam back to a camera. The laser and the camera are synchronized at a rate of 1KHz using a waveform generator.

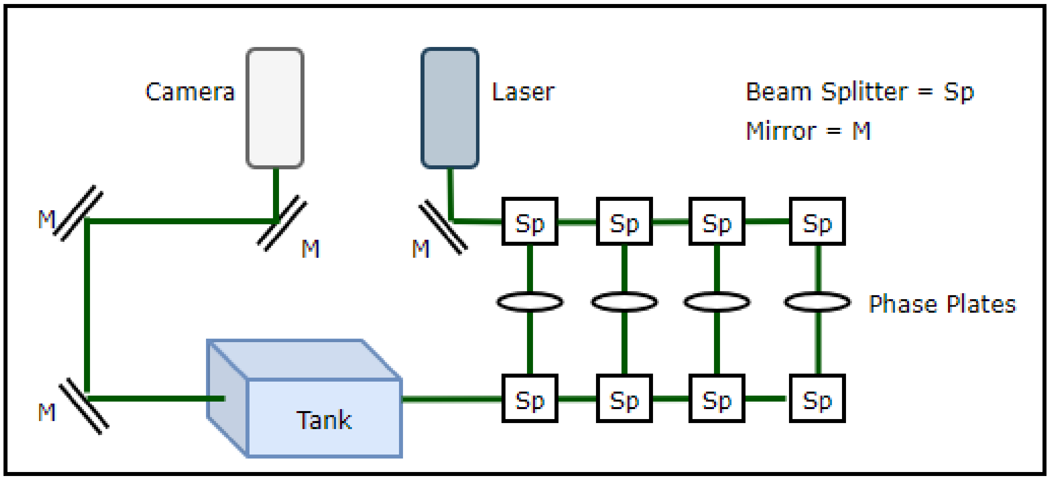


Fig. 1.1: OAM Benchtop Diagram.

An example of a few of the patterns created through a combination of four different phase plates is shown in Figure 1.2.

Turbulence causes intensity displacement (13), thereby imparting distortions in the intensity patterns at the receiver. This amounts to crosstalk (14; 15) between multiplexed

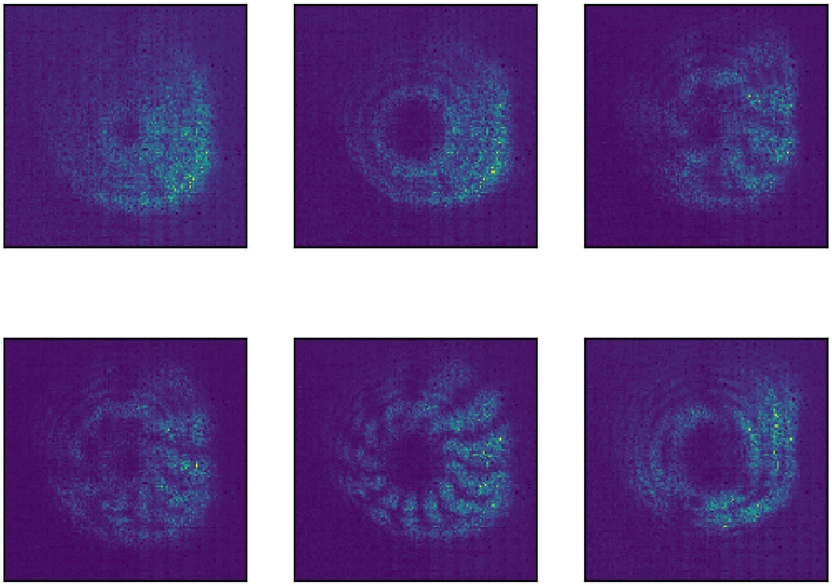


Fig. 1.2: Underwater OAM Patterns.

channels. Several approaches have been explored to address this issue in the literature. One approach (16) uses a beam to determine the amount of distortion present, then applies that information to correct subsequent transmissions. Other approaches (9; 17; 18; 19) use classification algorithms to recognize the modes. Ref. (20) used a novel approach applying a radon-cumulative distribution transform (R-CDT) of the received image. They found that detection algorithms trained on patterns in the R-CDT space did as well as or better than those based purely on the original OAM pattern images.

Refs. (9; 20; 21) applied deep learning to OAM free-space communications as a way to classify multiplexed OAM modes. This data included sets created from four levels of simulated turbulence. Both used a custom shallow CNN as well as AlexNet (22) for OAM mode classification.

A number of important questions remain unanswered from their work. Questions include, do their findings extend to the underwater domain? Where there was turbulence in free-space, do similar methods extend to attenuation in underwater communications?

Another factor to consider is classifier robustness when presented with data degraded beyond that which was in the training set. Given the established methods presented in (9) and (20), how do those approaches fare under heavy attenuation and turbulence? Is it possible to improve classifier robustness using models? Their work didn't explore these conditions and questions.

Additionally, if the shallow network and deep CNN (AlexNet) were able to get good results, would one of the more recent state of the art architectures provide greater performance? Are some architectures more robust in the face of turbulence or attenuation outside of what was present in the training set? As the end goal of this research is communications, classification robustness in the face of signal degradation is of primary importance.

1.2 Solution Overview

To address the questions and explore the problems outlined in Section 1.1, a number of principles and approaches are proposed. Robust OAM communications in underwater optical communications are explored and improved by applying principles of automatic differentiation, deep convolutional neural network architectures, optimal transport, and the R-CDT.

The area of automatic differentiation (AD) is an active area of research. Ref. (23) did a great survey on how AD works and the many ways it can be applied. While AD is used in backpropagation, there are many additional ways it can be used. For example, partial differentiation equation research has used AD in the formation of their models (24; 25). It has also been used in computer vision (26; 27), natural language processing (28; 29), optimization (30; 31; 32), and hyperparameter tuning (33; 34), to name a few. Given the flexibility of AD, can it be applied in the area of underwater OAM communications?

Since the success of AlexNet, CNNs have enjoyed a great deal of attention. The ImageNet (22) competition provides a forum for competing and extending the state of the art in image detection. Some of the landmark architectures include AlexNet (22), VGGNet (35), Inception (36; 37), ResNet (38), XCeption (39), ResNeXt (40), and DenseNet (41).

While these architectures provided state of the art performance on data sets like ImageNet, the question is whether those improvements extend to other image domains? Ref. (42) recently identified issues with some of the state of the art CNNs in terms of their ability to generalize. In light of this, does grabbing the latest and greatest CNN guarantee that optimal classification accuracy will be obtained?

While neural networks have been able to achieve impressive results in many different applications, one concern with using them is that their predictions may be inconsistent with physical realities. An area of interest in machine learning, and other areas of domain specific research, is creation of models using physics-based approaches. For example, (25; 43) applied AD to learning models for partial differential equations (PDE). They approached the question of models from the perspective of having large data quantities and working backwards toward finding models that uncovered the underlying physics of a system. Ref. (44) demonstrated a physics-guided neural network that was used for modeling lake temperatures. During training, it was shown that the neural network could predict values inconsistent with reality. By including physics models in the training loop, they were able to speed up and improve the model's results by using physics models to impose limits on realistic outputs.

Inclusion of physics-based models has the potential to increase CNN robustness in OAM pattern classification. Optimal transport is a mathematically rich, physics-based approach to representing an optimal path for the transfer of material from one location to another. Ref. (13) showed that the physics described by optimal transport can be adapted to electromagnetic waves. Given a source image and a received image, a transform can be created that captures how the signal moved through space. The optimal transport can be efficiently computed using the R-CDT (45). More in-depth information about optimal transport and the R-CDT can be found in Appendix C.

Optimal transport is used to learn a physics-based attenuation model in Chapter 3. An advantage of using the R-CDT is that it is able to capture the transform using a minimal number of samples (45). The R-CDT is also used in Chapter 2 as an alternative to standard approaches to using images in a CNN image classifier.

Given these examples of using physics-inspired training, is it possible to improve OAM mode classifier performance by incorporating these principles? Additionally, are there new ways to apply the R-CDT to allow improvement of OAM classifier robustness?

1.3 Research Questions and Contributions

While initial OAM data rate results are very exciting, a number of significant hurdles remain before it becomes a viable option for general use. Some factors include dealing with both turbulence (11) and signal attenuation (12). In looking at the current literature, it's apparent that there are many questions that remain to be answered. The work presented in subsequent sections address the specific interest of applying ML principles to improve OAM communications in underwater and free-space environments.

Chapter 2 explores whether current, proven methods in free-space optical communications, using the R-CDT, extend to underwater communications. Understanding of underwater OAM communication in attenuated environments is advanced through this research.

Chapter 3 explores model creation through training attenuation models from a custom CNN as well as the Radon-cumulative distribution transform. Questions addressed include exploring how attenuation can degrade classifier performance. Can physics-inspired models be created to improve classifier performance? Implications of using optimal transport and the R-CDT in this way are discussed, especially in the context of one-shot and few-shot learning. Attenuation models are developed using automatic differentiation.

Chapter 4 explores the current state of the art in convolutional neural networks in free-space and underwater environments. Do these newer CNNs produce corresponding improvements in OAM classification improvements? How well do the architectures perform when presented with images attenuated beyond that which was present in the training set? Additionally, some recent architectures have been optimized for resource constrained systems. Best performing architectures for resource constrained and resource rich systems are detailed.

Chapter 5 summarizes findings from the research.

This dissertation presents and demonstrates methods and principles for improving the state of the art in underwater optical communications under the presence of signal attenuation and turbulence. This is accomplished through advancing the state of the art in machine learning.

REFERENCES

[1] Statista, "Global internet use," 2020, data retrieved from statista site, https://www.statista.com/statistics/617136/digital-population-worldwide/.

[2] Y. Zhao, J. Li, X. Zhong, and H. Shi, "Physical-layer security in fractional orbital angular momentum multiplexing under atmospheric turbulence channel," *Optics Express*, vol. 27, no. 17, pp. 23 751–23 762, 2019. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-17-23751

[3] A. K. Majumdar and J. C. Ricklin, Eds., *Free-Space Laser Communications: Principles and Advances*, ser. Optical and Fiber Communications Reports. Springer-Verlag, 2008. [Online]. Available: https://www.springer.com/gp/book/9780387286525

[4] L. Allen, M. Beijersbergen, R. Spreeuw, and J. Woerdman, "Orbital angular momentum of light and transformation of laguerre gaussian laser modes," *Physical review. A*, vol. 45, pp. 8185–8189, 07 1992.

[5] J. Wang, J.-Y. Yang, I. Fazal, N. Ahmed, Y. Yan, H. HUANG, Y. Ren, Y. Yue, S. Dolinar, M. Tur, and A. Willner, "Terabit free-space data transmission employing orbital angular momentum multiplexing," *Nature Photonics*, vol. 6, pp. 488–496, 07 2012.

[6] G. Gibson, J. Courtial, M. J. Padgett, M. Vasnetsov, V. Pas'ko, S. M. Barnett, and S. Franke-Arnold, "Free-space information transfer using light beams carrying orbital angular momentum," *Optics Express*, vol. 12, no. 22, pp. 5448–5456. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-12-22-5448

[7] R. Chen, H. Zhou, M. Moretti, X. Wang, and J. Li, "Orbital angular momentum waves: Generation, detection and emerging applications," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[8] A. E. Willner, Z. Zhao, Y. Ren, L. Li, G. Xie, H. Song, C. Liu, R. Zhang, C. Bao, and K. Pang, "Underwater optical communications using orbital angular momentum-based spatial division multiplexing," *Optics Communications*, vol. 408,

pp. 21–25, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401817306818

[9] T. Doster and A. T. Watnik, "Machine learning approach to OAM beam demultiplexing via convolutional neural networks," *Applied Optics*, vol. 56, no. 12, pp. 3386–3396, 2017.

[10] J. Baghdady, K. Miller, K. Morgan, M. Byrd, S. Osler, R. Ragusa, W. Li, B. M. Cochenour, and E. G. Johnson, "Multi-gigabit/s underwater optical communication link using orbital angular momentum multiplexing," *Optics Express*, vol. 24, no. 9, pp. 9794–9805, May 2016. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-9-9794

[11] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde, "Transport-based model for turbulence-corrupted imagery," *Applied Optics*, vol. 57, no. 16, pp. 4524–4536, 2018. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-57-16-4524

[12] B. Cochenour, K. Dunn, A. Laux, and L. Mullen, "Experimental measurements of the magnitude and phase response of high-frequency modulated light underwater," *Applied Optics*, vol. 56, no. 14, pp. 4019–4024, 2017. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-56-14-4019

[13] J. M. Nichols, A. T. Watnik, T. Doster, S. Park, A. Kanaev, L. Cattell, and G. K. Rohde, "An optimal transport model for imaging in atmospheric turbulence," 2017, arXiv:1705.01050.

[14] J. A. Anguita, M. A. Neifeld, and B. V. Vasic, "Turbulence-induced channel crosstalk in an orbital angular momentum-multiplexed free-space optical link," *Applied Optics*, vol. 47, no. 13, pp. 2414–2429, May 2008. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-47-13-2414

[15] W. Nelson, J. P. Palastro, C. C. Davis, and P. Sprangle, "Propagation of bessel and airy beams through atmospheric turbulence," *Journal of the Optical*

Society of America A, vol. 31, no. 3, pp. 603–609, Mar 2014. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-31-3-603

[16] Y. Ren, G. Xie, H. Huang, C. Bao, Y. Yan, N. Ahmed, M. P. J. Lavery, B. I. Erkmen, S. Dolinar, M. Tur, M. A. Neifeld, M. J. Padgett, R. W. Boyd, J. H. Shapiro, and A. E. Willner, "Adaptive optics compensation of multiple orbital angular momentum beams propagating through emulated atmospheric turbulence," *Optics Letters*, vol. 39, no. 10, pp. 2845–2848, 2014. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-39-10-2845

[17] M. Krenn, R. Fickler, M. Fink, J. Handsteiner, M. Malik, T. Scheidl, R. Ursin, and A. Zeilinger, "Communication with spatially modulated light through turbulent air across vienna," *New Journal of Physics*, vol. 16, no. 11, p. 113028, 2014. [Online]. Available: https://doi.org/10.1088%2F1367-2630%2F16%2F11%2F113028

[18] E. M. Knutson, S. Lohani, O. Danaci, S. D. Huver, and R. T. Glasser, "Deep learning as a tool to distinguish between high orbital angular momentum optical modes," in *Optics and Photonics for Information Processing X*, K. M. Iftekharuddin, A. A. S. Awwal, M. G. Vázquez, A. Márquez, and M. A. Matin, Eds., vol. 9970, International Society for Optics and Photonics. SPIE, 2016, pp. 236 – 242. [Online]. Available: https://doi.org/10.1117/12.2242115

[19] S. Li and J. Wang, "Adaptive free-space optical communications through turbulence using self-healing bessel beams," *Scientific Reports*, vol. 7, p. 43233, 2017.

[20] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, "Demultiplexing vortex modes in optical communications using transport-based pattern recognition," *Optics Express*, vol. 26, no. 4, pp. 4004–4022, 2018.

[21] T. Doster and A. T. Watnik, "Measuring multiplexed OAM modes with convolutional neural networks," in *Lasers Congress 2016 (ASSL, LSC, LAC) (2016), paper LTh3B.2*. Optical Society of America, 2016, p. LTh3B.2. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=LSC-2016-LTh3B.2

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386

[23] B. Günes, P. A, R. Andreyevich, and S. Mark, "Automatic differentiation in machine learning," *The Journal of Machine Learning Research*, 2017. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3122009.3242010

[24] RaissiMaziar, "Deep hidden physics models," *The Journal of Machine Learning Research*, 2018. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3291125.3291150

[25] M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, Mar. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021999117309014

[26] S. M. A. Eslami, N. Heess, T. Weber, Y. Tassa, D. Szepesvari, K. Kavukcuoglu, and G. E. Hinton, "Attend, infer, repeat: Fast scene understanding with generative models," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 3233–3241.

[27] F. Srajer, Z. Kukelova, and A. Fitzgibbon, "A benchmark of selected algorithmic differentiation tools on some problems in computer vision and machine learning," *Optimization Methods and Software*, vol. 33, no. 4-6, pp. 889–906, Nov. 2018. [Online]. Available: https://doi.org/10.1080/10556788.2018.1435651

[28] K. Gimpel, D. Das, and N. A. Smith, "Distributed asynchronous online learning for natural language processing," in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning.*, Jul. 2010, pp. 213–222. [Online]. Available: https://www.aclweb.org/anthology/W10-2925

[29] Y. Goldberg, "A primer on neural network models for natural language processing," *CoRR*, vol. abs/1510.00726, 2015. [Online]. Available: http://arxiv.org/abs/1510.00726

[30] A. G. Baydin, B. A. Pearlmutter, and J. M. Siskind, "Diffsharp: Automatic differentiation library," 2015, arXiv:1511.07727.

[31] T. Schaul, S. Zhang, and Y. LeCun, "No more pesky learning rates," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28. PMLR, pp. 343–351. [Online]. Available: http://proceedings.mlr.press/v28/schaul13.html

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[33] D. Maclaurin, D. Duvenaud, and R. P. Adams, "Gradient-based hyperparameter optimization through reversible learning," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, p. 2113–2122.

[34] A. G. Baydin, R. Cornish, D. M. Rubio, M. Schmidt, and F. Wood, "Online learning rate adaptation with hypergradient descent," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=BkrsAzWAb

[35] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Sep. 2014, arXiv: 1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556

[36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014. [Online]. Available: https://arxiv.org/abs/1409.4842v1

[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[38] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[39] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” 2016. [Online]. Available: https://arxiv.org/abs/1610.02357v3

[40] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987–5995.

[41] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[42] A. Azulay and Y. Weiss, “Why do deep convolutional networks generalize so poorly to small image transformations?” *CoRR*, vol. abs/1805.12177, 2018. [Online]. Available: http://arxiv.org/abs/1805.12177

[43] M. Raissi, “Deep hidden physics models: Deep learning of nonlinear partial differential equations,” *Journal Machine Learning Research*, vol. 19, no. 1, p. 932–955, Jan. 2018.

[44] A. Karpatne, W. Watkins, J. S. Read, and V. Kumar, “Physics-guided neural networks (pgnn): An application in lake temperature modeling,” *ArXiv*, vol. abs/1710.11431, 2017.

[45] S. Kolouri, S. Park, and G. Rohde, “The radon cumulative distribution transform and its application to image classification,” *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 920–934, 2016.

CHAPTER 2

TRANSPORT-BASED PATTERN RECOGNITION VS. DEEP NEURAL NETWORKS IN UNDERWATER OAM COMMUNICATIONS

Patrick L. Neary, [1][2] Jonathan M. Nichols, [3][4] Abbie T. Watnik, [3] K. Peter Judd, [3] James R. Lindle, [5] Gustavo K. Rohde, [6] and Nicholas S. Flann [1]

2.1 Abstract

Comparisons between machine learning and optimal transport-based approaches in classifying images, previously made in turbulent free-space environments, are extended to underwater orbital angular momentum (OAM) communications. Mathematics are derived that justify optimal transport for use in attenuated water environments. OAM pattern de-multiplexing is performed with both optimal transport and deep neural networks and compared to similar tests in free-space. In clear water and attenuated environments, results are shown to be similar to free space communications. Additionally, some of the complications introduced by signal attenuation are highlighted. The radon cumulative distribution transform (R-CDT) is applied to the OAM patterns and its classification results are compared to standard machine learning-based image classification accuracies. To perform classification on the original OAM patterns and the R-CDT, the Nearest Subspace algorithm, shallow convolutional neural network (CNN), and deep neural network are used. It is shown that optimal transport and the R-CDT provide better results than CNNs in OAM pattern classification in underwater environments.

---

[1] Department of Computer Science, Utah State University, Old Main Hill, Logan, UT 84322, USA

[2] Space Dynamics Laboratory, 1695 N Research Park Way, North Logan, UT 84341

[3] Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA

[4] Dr. Nichols was responsible for the derivation that accounts for attenuation in optimal transport in water (Section 2.3) in the text and the appendices.

[5] DCS Corporation, 6909 Metro Park Drive, Suite 500, Alexandria, VA 22310, USA

[6] Dept. of Electrical and Computer Engineering, University of Virginia, 415 Lane Road, Charlottesville, VA 22908, USA

2.2 Introduction

Digital communication demands are continually increasing. Currently it is estimated that only 58% of the world's population is connected to the Internet (1). As access continues to be extended to remote areas and usage continues to proliferate, demand for bandwidth will only increase. Difficulties exist in providing access to remote areas, such as costs for laying cables, and dealing with difficult obstacles (in water or land) when laying cables. Other concerns related to communications are the problems of security and dealing with eavesdropping with signals (2).

One approach to addressing these issues is to use free space optical (FSO) communications. FSO provides the benefit of bypassing costly cable installations through geographically difficult regions. Another benefit of FSO is the fact that it is difficult to intercept the communication stream without degrading the signal and alerting the system to tampering attempts. However, security is an ongoing area of interest and study with OAM communications (2).

OAM is a property of electromagnetic (EM) waves. As coherent light is created and transmitted, EM radiation propagates in plane waves. In 1992, Ref. (3) was able to show, that in addition to having forward momentum, Laguerre-Gaussian (LG) beams also display orbital angular momentum. Assuming that light propagates along the z-axis, Allen et al. showed that LG beams travel in a helical pattern about that axis. The LG equation contains an azimuthal dependency expressed as $exp(-i\ell\phi)$, where ℓ is called the topologically-charged. When ℓ is 0, the beam propagates in a standard planar wavefront. When $|\ell| > 0$, the EM wave experiences angular momentum and travels in a helically shaped wavefront and the radius of the EM wave increases with ℓ. The sign on ℓ determines whether the EM helix propagates in a left-handed or right-handed direction.

One of the compelling properties of OAM is that topologically charged modes are orthogonal to each other. Consequently, OAM modes are good candidates for representing independent patterns that can be easily multiplexed/de-multiplexed in a communications link. Methods for de-multiplexing OAM patterns include adaptive optics (4), conjugate mode sorting (5; 6), spiral fringe counting (7), optical transformers (8), Doppler effect measurements (9), dove prism interferometers (10), DSP-based MIMO channel equalizers (11), and machine learning (ML) approaches (12; 13).

In ideal transmission environments, the beams can be demultiplexed at the receiver without loss or crosstalk of any signals. Ref. (14) showed that by multiplexing four OAM modes together, they were able to achieve >100 Tera-bit/s data rates. This was accomplished over short distances in an ideal environment.

Communications in real-world environments are complicated by varying degrees of attenuation and or changes in the refractive index of the medium, i.e., $n(\vec{x}, z) \neq 1$. Attenuation is caused by particles that either absorb or deflect signal. Turbulence is a frequent cause of variations in the index of refraction and can result in a loss of orthogonality among the modes. Turbulence creates crosstalk between channels.

Several approaches have been explored to address these issues. One approach (4) uses a separate non-OAM beam to determine the amount of phase distortion present to precompensate transmission of the OAM signals. Other approaches use classification algorithms to recognize the modes (15; 16; 17; 18). Ref. (19) used a novel approach applying a Radon-cumulative distribution transform (R-CDT) of the received image. This transform takes the received OAM patterns and maps them to the R-CDT space. They found that detection algorithms trained on patterns in the R-CDT space did as well as or better than those based purely on the original OAM pattern images.

This work focuses on ML and optimal transport-based demultiplexing approaches that are based on the problem physics. These approaches help correctly classifying OAM modes in the presence of both attenuation and variations in the medium's optical properties.

The main contributions of this work are extending R-CDT detection in OAM communications from the free space domain to the underwater domain. In addition, the mathematics extending optimal transport to attenuated environments are derived. This work explores the use of the optimal transport model from FSO communications, extends it to underwater optical communications (UWOC), and demonstrates a similar improvement in detection performance. Another contribution in this work is in looking at signal attenuation in water as opposed to turbulence in free-space. The remainder of the paper proceeds as follows: background on OAM and optimal transport (Section 2.3), experiment setup (Section 2.4), results

(Section 2.5) and final conclusion (Section 2.6).

2.3 Background and Prior Art

2.3.1 Signal Propagation and Orbital Angular Momentum

The electric field associated with a linearly polarized, monochromatic beam propagating in the z-direction is typically modeled as $E(\vec{x}, z, t) = U(\vec{x}, z)e^{i(\omega t - k_0 z)}\hat{\vec{x}}$. Here $\omega = 2\pi c/\lambda = k_0 c$ is the temporal frequency of oscillation, c is the speed of light, λ is the wavelength, and k_0 is the wavenumber and the vector $\hat{\vec{x}}$ encodes the direction of polarization in the plane transverse to the direction of propagation i.e., $\vec{x} \equiv \{x, y\}$ in Cartesian coordinates. In a homogeneous, isotropic medium characterized by refractive index $n(\vec{x}, z)$, and assuming the rate of transverse variations are slow relative to speed of propagation the complex amplitude of this electric field can be shown to obey:

$$-i2k_0 \frac{\partial U(\vec{x}, z)}{\partial z} + \nabla_X^2 U(\vec{x}, z) + k_0^2[n^2(\vec{x}, z) - 1]U(\vec{x}, z) = 0 \tag{2.1}$$

where ∇_X^2 is the Laplacian, acting in the transverse plane only.

There are many solutions to Eq. (2.1), particularly in the so-called “free-space” situation where $n(\vec{x}, z) = 1$ and the last term in Eq. (2.1) vanishes (20). For example, a variety of circularly symmetric solutions have been found which exhibit the OAM properties including Bessel (21), Bessel-Gauss (22), Laguerre-Gauss (3), Hermite-Gauss (23), Ince-Gauss (24), and Mathieu-Gauss (25). The complex amplitude of Laguerre-Gaussian (LG) beams, for example, can be represented in cylindrical coordinates as

$$\begin{aligned} U_{\ell,m}(r_0, \phi, z) = A \left[\frac{W_0}{W(z)}\right] \left[\frac{\sqrt{2} r_0}{W(z)}\right]^{|\ell|} L_m^{|\ell|}\left(\frac{2r_0^2}{W^2(z)}\right) \\ \times \exp\left(-\frac{r_0^2}{W^2(z)}\right) \exp\left(ik\frac{r_0^2}{2R(z)} + i\ell\phi + i\zeta(z)\right) \end{aligned} \tag{2.2}$$

where $\zeta(z) = -(|\ell|+2m+1)\tan^{-1}(2z/kW_0^2)$ is the Gouy phase and $W(z) = W_0\sqrt{1+4z^2/k^2W_0^4}$ is the beam waist (26). The function $L_m^{|\ell|}(\cdot)$ denote the Laguerre polynomials.

Different OAM modes are defined by the the topological charge or mode number, ℓ. In practice, the mode number can be controlled through vortex phase plates or spatial light modulators.

2.3.2 Lagrangian Model and Problem Statement

In a prior work (27), Nichols et al. developed image models consistent with the physics expressed in Eq. (2.1) by first describing the field as the phasor $U(\vec{x}_0, z) \equiv \rho^{1/2}(\vec{x}_0, z)\exp[i\psi(\vec{x}_0, z)]$. Specifically they were able to show that substituting this description of the electric field into Eq. (2.1), and neglecting diffraction, the magnitude of the complex electric field $\rho \equiv U^*U$ is governed by a transport equation which can be written (see Appendix A for the model development) as

$$\rho(\vec{x}_z) = \rho(\vec{x}_0)\det{}^{-1}\left(J_f(\vec{x}_0, z)\right) \tag{2.3}$$

while the associated transverse phase gradient $\frac{1}{k_0}\nabla_X\psi(\vec{x}_0, z) \equiv d\vec{x}_z/dz$ is governed by

$$\frac{d^2\vec{x}_z}{dz^2} = \nabla_X\eta(\vec{x}_z) \tag{2.4}$$

where $\eta(\vec{x}_z) = \left(\frac{n^2(\vec{x}_z)-1}{2}\right)$. Eq. (2.3) is simply the continuity equation in Lagrangian coordinates $\vec{x}_z = \vec{f}(\vec{x}_0, z)$, expressed in terms of the transverse Jacobian of the function that defines those coordinates. In other words, the total magnitude of the electric field is conserved as it is transported over a distance z under the action of $\vec{f}(\vec{x}_0, z)$. This coordinate transformation is, in turn, the solution to Eq. (2.4) and is clearly governed by the refractive index of the medium (see Ref. (28) for details). In the communications application, the function $\vec{f}(\vec{x}_0, z)$ is a disruptive influence that prevents an accurate identification of the multiplexed pattern that was sent. This stems from the fact that $\eta(\vec{x}, z)$ is a realization of a random process consistent with probabilistic models of the refractive index fluctuations (e.g., Kolmogorov turbulence).

The classification problem is then as follows: given a received image $\rho^{(\ell)}(\vec{x}_0, Z)$ that has traveled a distance Z and been corrupted by the medium according to Eq. (2.3), identify the correct mode number $\ell \in [1, L]$ of the "clean" pattern $\rho^{(\ell)}(\vec{x}_0, 0)$ that was sent. This is precisely the problem considered in Ref. (29). In the cited work, a new classifier was developed and demonstrated effective in this setting, outperforming a number of standard machine learning and neural net algorithms. We will describe this approach in what follows and then apply the classifier to the underwater communications applications alluded to earlier.

2.3.3 Classification in the R-CDT Domain

Before describing the classifier, the mathematics underlying the one and two-dimensional cumulative distribution transforms (CDTs) (30; 31) are reviewed.

In one dimension, an alternative formulation of Eq. (2.3) is

$$\int_{-\infty}^{x_0} \rho(u, z)du = \int_{-\infty}^{f(x_0,z)} \rho(u, 0)du \tag{2.5}$$

which again states that the total intensity present at $z = 0$ remains unchanged under the coordinate transformation induced by the physics of the problem. The expression in Eq. (2.5) is closely related to the cumulative distribution transform (CDT) introduced in Ref. (30). For strictly positive signals, the CDT is defined in terms of a reference signal $s_0(y),\ y \in \Omega_s$ as

$$\int_{-\infty}^{\widehat{\rho}_Z(y)} \rho(u, Z)du = \int_{-\infty}^{y} s_0(u)du \tag{2.6}$$

Choosing the reference domain $\Omega_s = [0, 1]$ and $s_0(y) = 1/|\Omega_s| = 1$, it can be shown that $\widehat{\rho}_Z(y)$ becomes the inverse of the cumulative distribution of the recorded signal $\rho(x_0, Z)$. The CDT $\widehat{\rho}_Z(y)$ is an invertible transform defined on the reference signal domain and the pairing $\rho(x_0, Z);\ \widehat{\rho}_Z(y)$ is thus analogous to the more familiar Fourier Transform pair.

We can use the properties of the CDT to develop a measure of similarity between signals for classification purposes. We write

$$\|\widehat{\rho}_Z(y) - \widehat{\rho}_0(y)\|^2 = \int_{X_0} \left[f(x_0, Z) - x_0\right]^2 \rho_0(x_0)dx_0 \tag{2.7}$$

as the amount of "effort" it takes to transform $\rho(x_0, 0)$ into $\rho(x_0, Z)$. In fact, the coordinate transformation f that minimizes this distance was shown in Ref. (32) to coincide with the solution to the physical equations of motion described in Eq. (2.4). Thus, a physically meaningful criteria for classification judges signals to be from the same class if the Euclidean distance between their respective CDTs is minimized.

The CDT and the similarity metric in Eq. (2.7) was extended to two-dimensions in Ref. (31) with help from the Radon transform. The Radon transform of an image $\rho(\vec{x}_0, z) \in \Omega_s \subset \mathbb{R}^2$, is denoted $\widetilde{\rho_0} = \mathscr{R}(\rho)$, and defined by

$$\widetilde{\rho}_\theta(t, z) \quad = \quad \int_{\Omega_s} \rho(\vec{x}_0, z)\delta(t - \vec{x}_0 \cdot \vec{\xi}_\theta)d\vec{x}_0 \tag{2.8}$$

Here, t is the perpendicular distance of a line from the origin and $\vec{\xi}_\theta = [\cos(\theta), \sin(\theta)]^T$, where θ is the angle over which the projection is taken. This is an intensity preserving transform that also possesses a well-defined inverse (29). In the Radon domain, the two-dimensional R-CDT is defined with respect to the reference image $s_0(\vec{x})$ as

$$\int_{-\infty}^{\widehat{\rho}_z(t,\theta)} \widetilde{\rho}_\theta(u, z)du = \int_{-\infty}^{t} \widetilde{s}_\theta(u, 0)du, \quad \forall\theta \in [0, \pi] \tag{2.9}$$

which is also invertible. Therefore, if one applies the Radon transform to the received signal, the two-dimensional CDT is "split" into a computation of many one-dimensional CDTs, one for each θ in the range $(0, \pi)$.

The Radon transform has a similar influence on the model presented in Eq. (2.3). Applying the Radon transform, one has that for every θ,

$$\widetilde{\rho}_\theta(t, z) = \widetilde{\rho}_\theta(t, 0)\left|\frac{df_\theta(t, z)}{dt}\right|^{-1} \tag{2.10}$$

where the collection of 1D mappings f_θ can be related to f; i.e., the f_θ still captures the physics of the problem (see Appendix A).

Given these properties, we can now formulate our strategy for solving the classification problem. As in the 1D case presented in Eq. (2.7) we suggest the metric

$$\begin{aligned} d^2(\rho_0, \rho_Z) &= \int_0^{\pi} \int_{-\infty}^{\infty} \left[f_\theta(t, Z) - t\right]^2 \widetilde{\rho}_\theta(t, 0) dt d\theta \\ &= \|\widehat{\rho}_z(t, \theta) - \widehat{\rho}_0(t, \theta)\|_2^2. \end{aligned} \tag{2.11}$$

as the "cost" associated with transforming the signal $\rho(\vec{x}_0, 0)$ into $\rho(\vec{x}_0, Z)$ in the Radon domain. This cost is also known as the sliced Wasserstein metric (33). It has been shown that the mapping $\vec{f} \equiv \vec{x}_z$ that solves Eq. (2.4) is the minimizer of both Eq. (2.11) and the kinetic energy associated with propagation (28). The distance d between the R-CDTs of two images is therefore assumed to be minimal when the only difference between those images is the distorting influence of the problem physics. Put another way, in the R-CDT domain the corrupted pattern should be closest, in the Euclidean sense, to its pristine counterpart. Different patterns (sent/received) would require more than simply the physics of propagation to transform one into the other and hence will be further apart in CDT space. This is true since the physics obeys the principle of least action, an assumption that underlies the derivation of many physical systems.

Based on the metric suggested in Eq. (2.11), a classifier can be suggested based on the understanding that a generative model for the R-CDT of a certain known pattern for the k-th class, denoted as $\widetilde{\varphi}_\theta^{(k)}(t, z)$, is given by some unknown displacement (in Radon domain) applied to it via:

$$\frac{df_\theta(t, Z)}{dt} \widetilde{\varphi}^{(k)}[f_\theta(t, Z), \theta] \tag{2.12}$$

where $f_\theta(t, Z)$ is a model approximation for the unknown transport-based distortion imparted on the wave as it travels through a non-uniform medium. Using the composition property stated in Ref. (31), we have that the deformed pattern in R-CDT space can be expressed as $f_\theta^{-1}\left[\tilde{\varphi}_\theta^{(k)}(t, Z), Z\right]$. Furthermore, if we assume that $f_\theta^{-1}(\cdot, \cdot)$ forms a convex subspace within the set of all possible 1D diffeomorphisms, then an easy solution to the classification problem can be proposed. In Ref. (29), the authors propose a classification method that utilizes the convex subspace assumption above to perform image classification.

The principle is to estimate the subspace that models all possible variations of the observed deformed pattern $f_\theta^{-1}\left[\tilde{\varphi}_\theta^{(k)}(t, Z), Z\right]$ (in Radon space), by simply estimating a basis for each subspace formed by the training data in that class. The approach taken here is the same as the one proposed in Ref. (29) where the Principal Component Analysis (PCA) technique is used to obtain a basis for the training set of each class. The testing procedure then simply consists of transforming the input test data, computing the least squares distance to each of the trained subspaces, and assessing the class of the test sample to be the same of the closest R-CDT subspace.

In summary, for data modeled by Eq. (2.3), minimizing the Euclidean norm in the R-CDT domain will give the pattern that required the “least action” to transform into the received image. Thus, one might expect good classification performance in situations such as optical communications where we have just shown the observed data indeed obey such a model.

2.4 Experiment Setup

2.4.1 Hardware

The laser source used in these measurements was a diode-pumped solid state laser (Bright Solutions ONDA 532) that operates at 532 nm and produces 5 nS pulses with 300 uJ/pulse. The laser is externally triggered at 1 kHz using a waveform generator (Agilent 33600A) and the output beam was expanded to fill the aperture of the static vortex phase plates. The output beam was then split into four equal intensity beams and each beam was normally incident on one of the four phase plates of different charge [1, 4, -6, -8], imprinting a different OAM phase on each beam. The phase plates were fabricated at Clemson University. The beams were then coherently recombined using beamsplitters and reflected through a water tank 1.2 m in length. The transmitted intensity patterns were captured by a high performance, fast-frame-rate camera (Photron FASTCAM SA-Z) that was externally triggered at the rate of 1000 frames per second, synchronized with the laser pulses.

The 16 OAM mode patterns were manually selected by blocking one or more of the beams.

The camera spatial resolution is 1024x1024 pixels with a 20 m pitch and resolution depth of 12 bits. Five micron polyamid seeding particles (Dantec Dynamics) are added to the water tank thereby introducing scattering of the beam, resulting in beam attenuation in the forward direction. Beam attenuation is continuously monitored by measuring the transmission of a 15 mW, 532 nm CW laser beam propagating through the tank parallel to the OAM beam. Small pumps agitate the water to ensure that the particles remain in suspension and homogeneously distributed.

In previous work ferroelectric spatial light modulators (SLMs) were used for generating different phase patterns. Since SLMs are relatively slow, specially fabricated phase plates were created to improve throughput rates, to simplify configuration, and to prepare for eventual use in real environments.

2.4.2 Data Set

The influence of an attenuating medium on the propagation physics are discussed in Appendix A. In summary, attenuation by scattering or absorption is modeled here by the complex refractive index $n(\mathbf{x}) - i\kappa$ where κ is the well-known "extinction coefficient". This results in the removal of signal intensity as the beam moves in the direction of propagation. In this work, it is assumed that this effect is uniform in the transverse dimension (i.e., κ is not a function of $\vec{x}_z$). The resulting signal model is therefore

$$\rho(\vec{x}_z) = \rho(\vec{x}_0)\text{det}^{-1}\left[J_f(\vec{x}_0, z)\right] e^{-2\kappa k z}. \tag{2.13}$$

As expected, as the beam propagates, it loses photons at an exponential rate governed by the attenuation length (AL) $(2\kappa k)^{-1}$. The end result of these losses is a lower signal-to-noise ratio (SNR) for the received signal $\rho(\vec{x}_z)$. Prior to computing the R-CDT, all images are first normalized to the same intensity to enforce the continuity constraint. Those images collected at longer attenuation lengths will therefore have their SNR lowered by the exponential factor in Eq. (2.13).

One thousand images from each of the 16 permutations of the OAM beam set were captured for the analysis, resulting in 16,000 images per data set. The base attenuation length, 0, is obtained in clear water. The subsequent attenuation lengths are 4, 8, and 12. The data sets collected at these attenuation lengths are referred to as AL0, AL4, AL8, and AL12.

The AL sets are further combined to create data sets with mixed attenuation lengths. The combined sets are referred to as AL0-4, AL0-4-8, and AL0-4-8-12. In creating combined data sets, the total data set size is kept to approximately the same size as the originals, i.e. a total set size of 16,000 or 1,000 for each pattern. In AL0-4, for example, AL0 and AL4 beam patterns are randomly sampled where 50% come from AL0 and 50% come from AL4. This allows computational complexity between training sets to be kept equal.

After combined data sets are created, each data set is randomly sampled and split into 70%/15%/15% for training, validation, and testing. Figure 2.1 shows six OAM modes with images taken from each attenuation level.

The images are cropped and downsampled to capture the patterns of interest and exclude borders that do not contain relevant information, thereby reducing computation time. The clipped images are 512x512 in size. The images are further downsampled to 128x128 and are used as the basis for the experiments performed in this paper. Figure 2.2 contains example images of the 15 multiplexed OAM beam patterns sampled from the AL0 set.

Two image spaces are used in subsequent experiments, the downsampled images and their transform. A Radon-cumulative distribution transform is performed on each of the 128x128 AL data sets to create the second image space. An example of an OAM mode and its corresponding R-CDT is shown in Figure 2.3.

For the R-CDT set creation, the R-CDT transform was run between 0-90 degrees with a step size of 2 degrees on the OAM images. In computing the R-CDT, a mapping from one image to another is performed. A base image with dimensions equal to the original is created that has uniform intensity, where all pixel intensities sum to one.

2.4.3 Experiment descriptions

An objective of this work is to see if OAM pattern classification in an underwater environment, subjected to attenuation, is similar to the free space results of Ref. (19) under

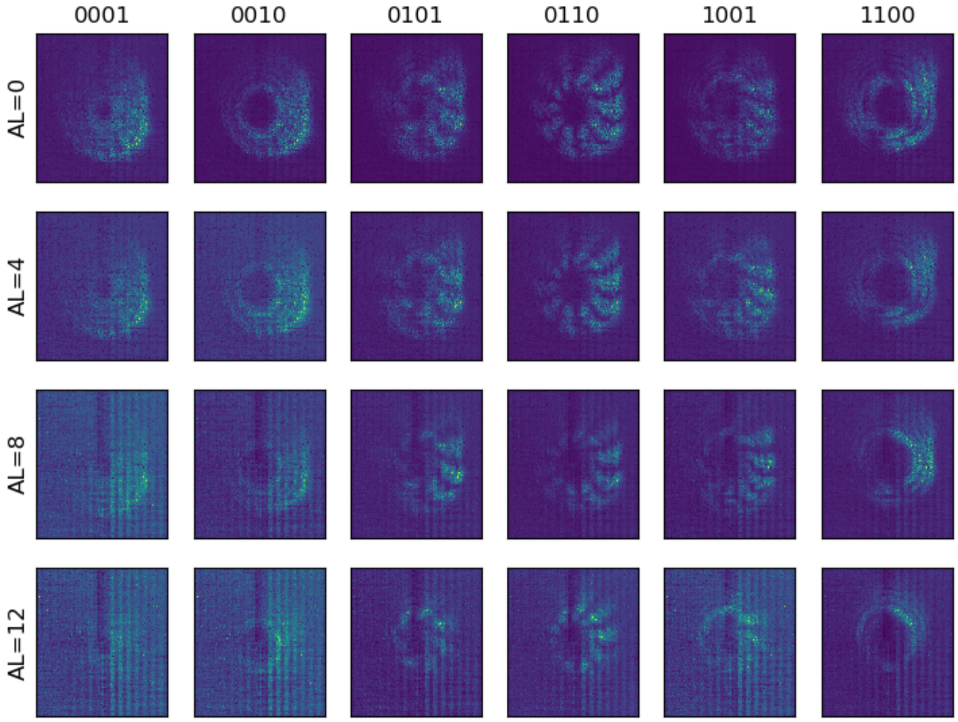


Fig. 2.1: Examples of attenuated images from the underwater environment. Each row represents a specific level of attenuation. The first row contains attenuation level 0. The subsequent rows show progressively higher levels of attenuation. Each column contains a specific OAM beam pattern.

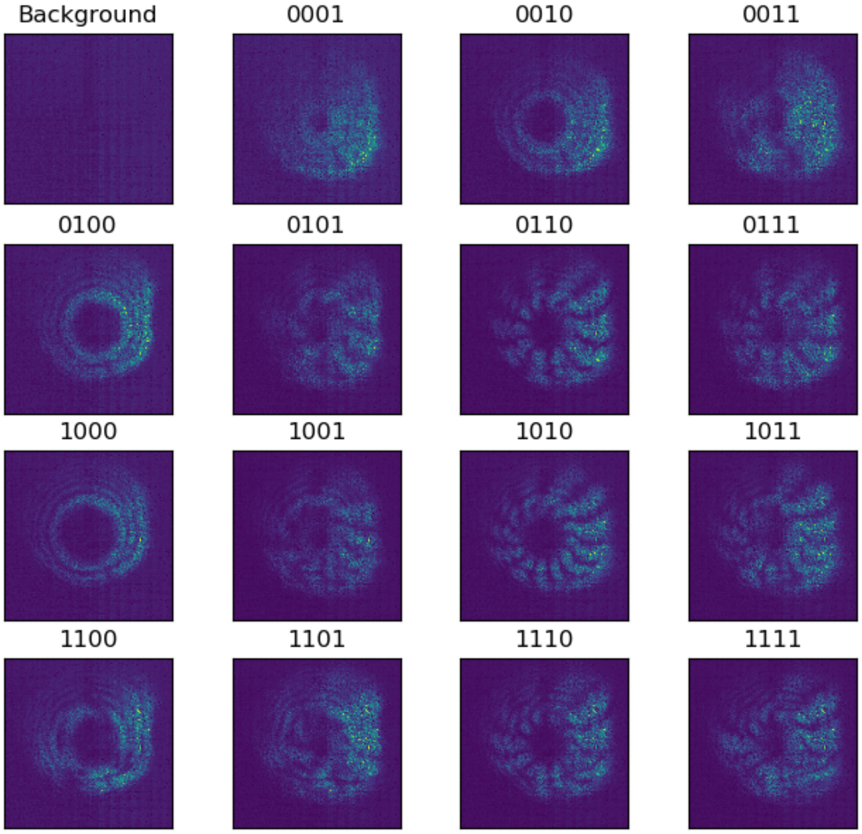


Fig. 2.2: OAM modes used for underwater optical communications. The background image is the image taken without the laser being active. Images are individual samples from the AL0 set.

turbulence. An additional goal is to identify how attenuation affects the performance of OAM beam classification and which, if any, of the algorithms provides the most robust performance.

Ref. (19) established that linear discriminant analysis (LDA) with R-CDT provides superior computational training speed and accuracy compared to several deep learning architectures. Recently Ref. (29) provided an updated approach to computing the R-CDT and provides a Python package to implement it easily. The software is available on GitHub (34). The package also provides a Nearest Subspace (NS) algorithm for performing classification. The NS algorithm estimates the subspace for possible variations in the R-CDT patterns and performs classification based on those subspace estimates (29).

Following the initial approach of Ref. (19), LDA classification was performed on the PCA of 4000 components of features found within the images. Accuracy results and computational speed were then compared to the classifier recently provided by Ref. (29) and an improvement in both were observed. Consequently, the NS classification algorithm is used in this work instead of LDA.

Similar to Ref. (19), this work contrasts the NS algorithm (in place of LDA) and two convolutional neural networks (CNN). For these experiments, images and their corresponding R-CDT are compared against each other to identify which provides better discrimination.

The shallow CNN consists of a CNN layer with sixteen 11x11 filters configured with a stride of 3x3. Next is a CNN layer with thirty-two 3x3 filters with a stride of 3x3. These layers are followed by a max pool layer, ReLU, and a fully connected layer. Weights in this shallow CNN are initialized according to Ref. (35). Adam is the selected optimizer (36). The deep neural network selected is AlexNet (37) for comparison with Ref. (19). When AlexNet and the shallow CNN are used with 128x128 images, the input kernels are 11x11. When the R-CDT sets are used, the kernels dimensions are configured as 11x5.

2.5 Results

This section presents a comparison of R-CDT and image space classification accuracies. The effects of attenuation on classification performance are also presented.

Table 2.1 shows the side-by-side classification results of R-CDT and OAM image pattern prediction accuracy for three different classification algorithms. The first column shows the training sets that are formed from various combinations of AL0-AL12. The second column shows the NS-based R-CDT results. Columns three and four show results for the Shallow CNN and columns five and six show results for the Deep CNN. The Shallow and Deep CNNs were trained on both the R-CDT image sets as well as the OAM images.

As displayed in column 2, the NS-based R-CDT algorithm outperforms the CNN methods on all combinations of the data sets. The Shallow and Deep CNN columns show that accuracies using the R-CDT patterns perform as well as or better than the image-based classifier the majority of the time.

The following test explores and contrasts how well the algorithms perform with limited numbers of training samples. Recently, Ref. (29) improved the R-CDT using the NS algo-

Table 2.1: **R-CDT to image space comparison across classifiers**

	NS	Shallow CNN		Deep CNN	
AL	R-CDT	R-CDT	Img	R-CDT	Img
0	100	100	99.5	99.8	99.4
4	100	99.2	99.6	99.5	99.7
8	100	99.8	99.3	99.4	99.6
12	100	99.9	98.8	99.7	98.4
0-4	100	100	99.4	99.6	99.2
0-4-8	100	99.9	99.4	99.2	99.4
0-4-8-12	100	99.9	99.0	99.0	99.7

rithm. These tests compare accuracies of the NS algorithm trained with R-CDT patterns, a shallow CNN trained with images, and a deep CNN named VGG-11 ((38)) trained with images. The number of training samples per class are from the set {1, 2, 4, 8, 16, 32, 64, 128, 256, 512}. The previous tests used approximately 700 images per class for training.

Figures 2.4 and 2.5 provide curves plotting classification accuracy of each algorithm given a fixed number of samples per class. Both figures show that the R-CDT combined with the Nearest Subspace algorithm is able to quickly learn distinguishing features with a limited number of training examples per class. Results for the other data sets follow the trends presented for AL0 and AL4 as shown in Figures 2.4 and 2.5.

Table 2.1 shows that the NS algorithm coupled with the R-CDT provides the best results. Each CNN architecture is trained and evaluated with R-CDT patterns. The CNN is then trained and evaluated with images. It is found that the R-CDT, in general, works as well as or outperforms the images based classification.

Looking closely at Table 2.1, the R-CDT-based test accuracies are consistently better than OAM image-based test accuracies for both the NS and CNN classification algorithms. Intuitively, it makes sense for the R-CDT space to provide better results than the original OAM image space. By performing the R-CDT, the images are moved into a new space where there is more separation between the classes, as shown in Ref. (19). Because there is greater separation between the classes in the R-CDT space, the decision boundaries that the classification algorithms learn are easier to differentiate and classification results are improved over what is achieved in the image space.

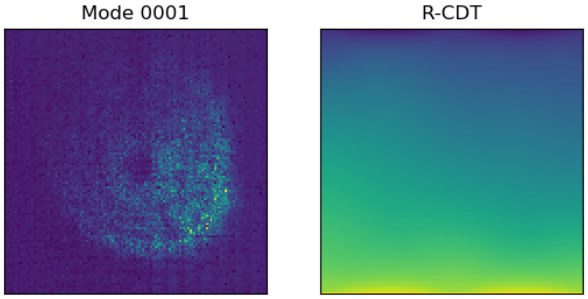


Fig. 2.3: Example of OAM pattern mode 0001 and its corresponding R-CDT.

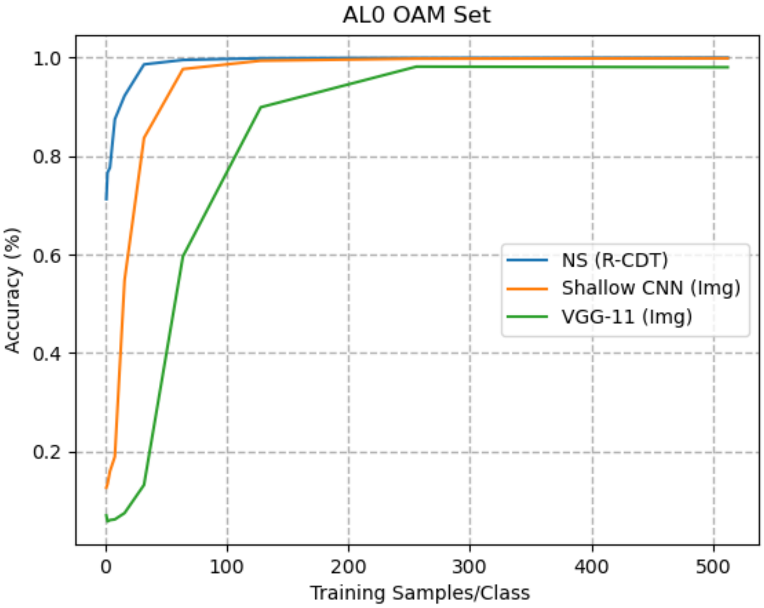


Fig. 2.4: Accuracy curves for AL0 data set.

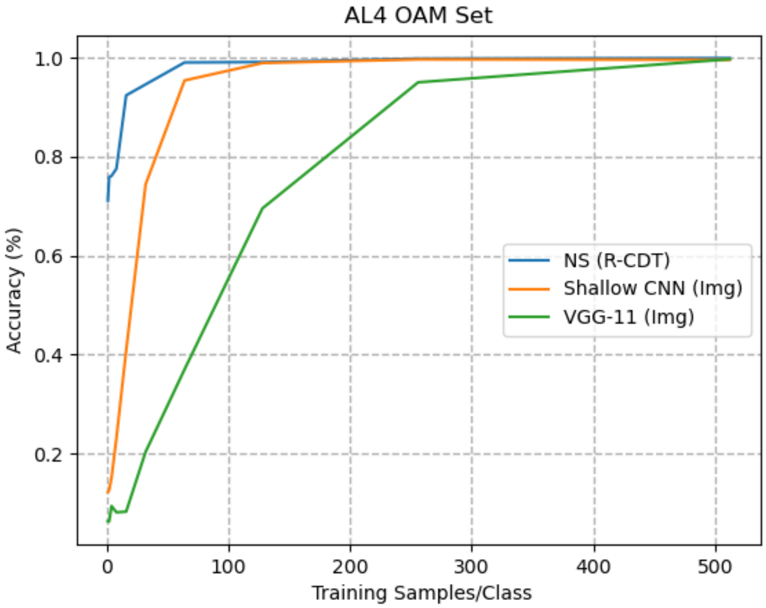


Fig. 2.5: Accuracy curves for AL4 data set.

Figures 2.4 and 2.5 display similar characteristics to those presented in Ref. (29). The NS algorithm, combined with the R-CDT, obtains better results than CNNs given the same number of training samples per class.

2.6 Summary and Conclusions

OAM beam patterns were collected at four different attenuation levels: AL0, AL4, AL8, and AL12. From these data sets, three new composite sets were created by combining different attenuation levels. These composite data sets are AL0-4, AL0-4-8, and AL0-4-8-12.

These data sets were used to evaluate how well training in the R-CDT space compared to training in image space. These sets were also used to explore how well different classification algorithms perform relative to each other. Concretely, the algorithms selected were the NS R-CDT classifier, a shallow convolutional neural network, and a deep CNN.

This work first set out to justify the use of transport theory in attenuated underwater environments. We then set out to see if classification gains seen by the R-CDT over image patterns in free space, extended to attenuated environments in water. It was found that, in the underwater environment, the R-CDT space performed better than the image space. This outcome correlates with what was observed in free space as reported in Ref. (19) and in Ref. (29).

Several areas exist for further exploration. In this work, a negligible amount of turbulence was present in the underwater environment, so a direct study of turbulence in water versus free-space (as presented in (19)) was not possible. Future work can perform a direct comparison of turbulence in water to free-space turbulence. Additionally, in future work it will be important to allow both turbulence and attenuation to vary simultaneously, to gain a full picture of the best algorithms for classification. Another interesting direction of research is to explore other transforms, to see what additional gains can be made in classification accuracy in the presence of signal degrading conditions.

2.7 Disclosures

The authors declare no conflicts of interest.

2.8 Acknowledgements

We would like to acknowledge support for this project from the US Naval Research Laboratory through a 6.1 base program for orbital angular momentum and a 6.2 base program for underwater communications. Thank you to the research group of Eric Johnson for fabrication of the vortex phase plates.

2.9 Appendix A: Image Model in Lagrangian Coordinates for Attenuated Beams

Representing the complex amplitude as $U(\vec{x}_0, z) = \rho^{1/2}(\vec{x}_0, z)e^{i[\psi(\vec{x}_0,z)]}$ and substituting into Eq. (2.1) yields a complex equation in ρ, ψ. The imaginary and real portions are, respectively,

$$\frac{\partial \rho(\vec{x}_0, z)}{\partial z} + \nabla_X \cdot [\rho(\vec{x}_0, z)\vec{u}(\vec{x}_0, z)] = 0 \tag{2.14a}$$

$$\frac{d\vec{u}(\vec{x}_0, z)}{dz} + [\vec{u}(\vec{x}_0, z) \cdot \nabla_X]\vec{u}(\vec{x}_0, z) = \nabla_X \eta(\vec{x}_z) + \frac{1}{2k_0^2}\frac{\nabla_X^2 \rho^{1/2}(\vec{x}_0, z)}{\rho^{1/2}(\vec{x}_0, z)} \tag{2.14b}$$

where $\vec{u}(\vec{x}_0, z) \equiv \frac{1}{k_0}\nabla_X \psi(\vec{x}_0, z)$. Expression (2.14a) governs the magnitude of the electric field and is the familiar continuity equation. The second expression is a conservation of momentum where the phase gradient acts as a dimensionless "velocity" and governs the change in optical path in the transverse dimension per unit change in the direction of propagation. The second term on the right hand side of Eq. (2.14b) is of negligible magnitude (note k_0^{-2} dependence) and can be safely neglected (39).

Both expressions can be simplified to some extent with a change to Lagrangian coordinates. Let $\vec{x}_z \equiv \vec{f}(\vec{x}_0, z)$ be a vector function of the coordinates $\vec{x}_0, z$. In this coordinate system, material derivatives, e.g., $\partial_z \rho(\vec{x}_0, z) + \nabla_X \rho(\vec{x}_0, z)\vec{u}(\vec{x}_0, z)$, become ordinary derivatives $d\rho(\vec{x}_z)/dz$. The velocity of the field then becomes simply $\vec{u}(\vec{x}_0, z) = d\vec{x}_z/dz$. In this formulation the expressions in Eq. (2.14) become

$$\frac{d\rho(\vec{x}_z)}{dz} + \rho(\vec{x}_z)\left[\nabla_X \cdot \vec{u}(\vec{x}_z)\right] = 0 \tag{2.15a}$$

$$\frac{d^2\vec{u}(\vec{x}_z)}{dz^2} = \nabla_X \eta(\vec{x}_z). \tag{2.15b}$$

Equation (2.15a) is now an ordinary differential equation in intensity (40) which can be solved to yield

$$\rho(\vec{x}_z) = \rho(\vec{x}_0)\exp\left[-\int_0^z \nabla_X \cdot \vec{u}(\vec{x}_s)ds\right] \tag{2.16}$$

To arrive at our data model described in Eq. (2.3) we require a final identity. Start with the expression that defines the Lagrangian coordinates $d\vec{f}(\vec{x}_0, z)/dz = \vec{u}(\vec{x}_z)$. Take the spatial gradient $\nabla_{\vec{x}_0}$ of both sides, and simplifying (re-arranging differential elements) gives

$$\frac{dJ_f(\vec{x}_0, z)}{dz} = \left[\nabla_X \cdot \vec{u}(\vec{x}_z)\right] J_f(\vec{x}_0, z). \tag{2.17}$$

where $J_f(\vec{x}_0, z)$ is the 2×2 Jacobian of the Lagrangian vector function $\vec{f}(\vec{x}_0, z)$. Taking the determinant of both sides then results in an ordinary differential equation in $\det[J_f(\vec{x}_0, z)]$.

Solving this equation (noting $\det[J_f(\vec{x}_0, 0)] = 1$) yields the identity (40; 41)

$$\det[J_f(\vec{x}_0, z)] = \exp\left(\int_0^z \nabla_X \cdot \vec{u}(\vec{x}_s) ds\right). \tag{2.18}$$

That is to say, the determinant of the Jacobian of the vector function governing the optical path is directly related to the path integral of the divergence. Substituting Eq. (2.18) into Eq. (2.16) then yields our image model presented in Eq. (2.3).

The above analysis presumes zero losses during signal propagation. However, assuming a non-zero extinction coefficient, the refractive index can be written $n(\vec{x}, z) - i\kappa$ with extinction coefficient κ. The governing equations, in Lagrangian coordinates, then become

$$\frac{d\rho(\vec{x}_z)}{dz} + \rho(\vec{x}_z)\left[\nabla_X \cdot \vec{u}(\vec{x}_z)\right] = -2\kappa n(\vec{x}_z) k_0 \rho(\vec{x}_z) \tag{2.19a}$$

$$\frac{d^2\vec{u}(\vec{x}_z)}{dz^2} = \nabla_X \tilde{\eta}(\vec{x}_z). \tag{2.19b}$$

where $\tilde{\eta}(\vec{x}_z) = [n^2(\vec{x}_z) - \kappa^2 - 1]/2$. As expected, non-zero losses results in a "sink" term in the continuity equation. The image model in Lagrangian coordinates therefore reads

$$\rho(\vec{x}_z) = \rho(\vec{x}_0) \exp\left[-\int_0^z \nabla_X \cdot \vec{u}(\vec{x}_s) ds\right] \exp[-2\kappa n(\vec{x}_z) k_0 z] \tag{2.20a}$$

or

$$\rho(\vec{x}_z) = \rho(\vec{x}_0) \det^{-1}\left[J_f(\vec{x}_0, z)\right] \exp[-2\kappa n(\vec{x}_z) k_0 z]. \tag{2.20b}$$

While this model predicts that the energy loss will be spatially non-uniform, depending on the varying refractive index, these variations will be treated as uniform in our analysis of signal loss. That is to say, at a given propagation distance, the signal loss due to absorption is assumed uniform across the image and proportional to the exponential factor $\exp(-2\kappa k z)$ where $k = \bar{n} k_0$ is the (uniform) wavenumber implied by the average refractive index.

2.10 Appendix B: Relating $f_\theta(t)$ to $\vec{f}(\vec{x}_0, z)$

In this section, we relate the problem physics, as encoded in the Lagrangian mapping

$\vec{f}(\vec{x}_0, z)$, to the corresponding function(s) in the Radon Transform domain. Taking the Radon Transform of both sides of Eq. (2.14a) yields (42)

$$\frac{\partial \widetilde{\rho}_\theta(t,z)}{\partial t} + \frac{d}{dt}\left[\widetilde{\rho}_\theta(t,z) u_\theta(t,z)\right] = 0 \tag{2.21}$$

where

$$u_\theta(t,z) \equiv \frac{df_\theta(t,z)}{dt} = \frac{\mathscr{R}\left[\rho(\vec{x}_0,z)\vec{u}(\vec{x}_0,z)\cdot\vec{\xi}_\theta\right]}{\widetilde{\rho}_\theta(t,z)} \tag{2.22}$$

and $\vec{u}(\vec{x}_0, z) = d\vec{f}(\vec{x}_0, z)/dz$. The Radon transform therefore splits the continuity equation in the transverse plane into a collection of such equations, one for each angle θ. Moreover, the coordinate mapping in transform space is directly related to the Cartesian mapping that defines the problem physics via Eq. (2.22). The collection of one-dimensional functions f_θ therefore carries the same information about the problem physics as does the vector function $\vec{f}$. In the instance of spatially constant "velocity" movement of intensity, Eq. (2.22) implies quite sensibly that the mapping in Radon space is the projection of the Cartesian mapping onto the vector $\vec{\xi}_\theta$.

REFERENCES

[1] Statista, “Global internet use,” 2020, data retrieved from statista site, https://www.statista.com/statistics/617136/digital-population-worldwide/.

[2] Y. Zhao, J. Li, X. Zhong, and H. Shi, “Physical-layer security in fractional orbital angular momentum multiplexing under atmospheric turbulence channel,” *Optics Express*, vol. 27, no. 17, pp. 23 751–23 762, 2019. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-17-23751

[3] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, “Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes,” *Physical Review A*, vol. 45, no. 11, pp. 8185–8189, 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.45.8185

[4] Y. Ren, G. Xie, H. Huang, C. Bao, Y. Yan, N. Ahmed, M. P. J. Lavery, B. I. Erkmen, S. Dolinar, M. Tur, M. A. Neifeld, M. J. Padgett, R. W. Boyd, J. H. Shapiro, and A. E. Willner, “Adaptive optics compensation of multiple orbital angular momentum beams propagating through emulated atmospheric turbulence,” *Optics Letters*, vol. 39, no. 10, pp. 2845–2848, 2014. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-39-10-2845

[5] G. Gibson, J. Courtial, M. J. Padgett, M. Vasnetsov, V. Pas’ko, S. M. Barnett, and S. Franke-Arnold, “Free-space information transfer using light beams carrying orbital angular momentum,” *Optics Express*, vol. 12, no. 22, pp. 5448–5456, Nov 2004. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-12-22-5448

[6] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger, “Entanglement of the orbital angular momentum states of photons,” *Nature*, vol. 412, no. 6844, pp. 313–316, 2001. [Online]. Available: http://www.nature.com/articles/35085529

[7] M. S. Soskin, V. N. Gorshkov, M. V. Vasnetsov, J. T. Malos, and N. R. Heckenberg, “Topological charge and angular momentum of light beams carrying optical

vortices," *Physical Review A*, vol. 56, pp. 4064–4075, Nov 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.56.4064

[8] M. P. J. Lavery, G. C. G. Berkhout, J. Courtial, and M. J. Padgett, "Measurement of the light orbital angular momentum spectrum using an optical geometric transformation," *Journal of Optics*, vol. 13, no. 6, p. 064006, apr 2011. [Online]. Available: https://doi.org/10.1088%2F2040-8978%2F13%2F6%2F064006

[9] M. P. J. Lavery, F. C. Speirits, S. M. Barnett, and M. J. Padgett, "Detection of a spinning object using light's orbital angular momentum," *Science*, vol. 341, no. 6145, pp. 537–540, 2013. [Online]. Available: https://science.sciencemag.org/content/341/6145/537

[10] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial, "Measuring the orbital angular momentum of a single photon," *Physical Review Letters*, vol. 88, no. 25, p. 257901, 2002.

[11] Y. Ren, Z. Wang, G. Xie, L. Li, A. J. Willner, Y. Cao, Z. Zhao, Y. Yan, N. Ahmed, N. Ashrafi, S. Ashrafi, R. Bock, M. Tur, and A. E. Willner, "Atmospheric turbulence mitigation in an OAM-based MIMO free-space optical link using spatial diversity combined with MIMO equalization," *Optics Letters*, vol. 41, no. 11, pp. 2406–2409, 2016.

[12] X. Cui, X. Yin, H. Chang, H. Liao, X. Chen, X. Xin, and Y. Wang, "Experimental study of machine-learning-based orbital angular momentum shift keying decoders in optical underwater channels," *Optics Communications*, vol. 452, pp. 116–123, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003040181930608X

[13] S. Lohani, E. M. Knutson, M. O'Donnell, S. D. Huver, and R. T. Glasser, "On the use of deep neural networks in optical communications," *Applied Optics*, vol. 57, pp. 4180–4190, 2018. [Online]. Available: http://arxiv.org/abs/1806.06663

[14] H. Huang, G. Xie, Y. Yan, N. Ahmed, Y. Ren, Y. Yue, D. Rogawski, M. J. Willner, B. I. Erkmen, K. M. Birnbaum, S. J. Dolinar, M. P. J. Lavery, M. J. Padgett, M. Tur, and A. E. Willner, "100 tbit/s free-space data link enabled

by three-dimensional multiplexing of orbital angular momentum, polarization, and wavelength," *Optics Letters*, vol. 39, no. 2, pp. 197–200, 2014. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-39-2-197

[15] M. Krenn, R. Fickler, M. Fink, J. Handsteiner, M. Malik, T. Scheidl, R. Ursin, and A. Zeilinger, "Communication with spatially modulated light through turbulent air across vienna," *New Journal of Physics*, vol. 16, no. 11, p. 113028, 2014. [Online]. Available: https://doi.org/10.1088%2F1367-2630%2F16%2F11%2F113028

[16] E. M. Knutson, S. Lohani, O. Danaci, S. D. Huver, and R. T. Glasser, "Deep learning as a tool to distinguish between high orbital angular momentum optical modes," in *Optics and Photonics for Information Processing X*, K. M. Iftekharuddin, A. A. S. Awwal, M. G. Vázquez, A. Márquez, and M. A. Matin, Eds., vol. 9970, International Society for Optics and Photonics. SPIE, 2016, pp. 236 – 242. [Online]. Available: https://doi.org/10.1117/12.2242115

[17] T. Doster and A. T. Watnik, "Machine learning approach to OAM beam demultiplexing via convolutional neural networks," *Applied Optics*, vol. 56, no. 12, pp. 3386–3396, 2017.

[18] S. Li and J. Wang, "Adaptive free-space optical communications through turbulence using self-healing bessel beams," *Scientific Reports*, vol. 7, p. 43233, 2017.

[19] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, "De-multiplexing vortex modes in optical communications using transport-based pattern recognition," *Optics Express*, vol. 26, no. 4, pp. 4004–4022, 2018.

[20] U. Levy, S. A. Derevyanko, and Y. R. Silberberg, "Light modes of free space," *Progress in Optics*, vol. 61, pp. 237–281, 2016.

[21] J. Durnin, J. J. Miceli, and J. H. Eberly, "Diffraction-free beams," *Physical Review Letters*, vol. 58, no. 15, pp. 1499–1501, 1987. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.58.1499

[22] F. Gori, G. Guattari, and C. Padovani, "Bessel-gauss beams," *Optics Communications*, vol. 64, no. 6, pp. 491–495, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0030401887902768

[23] A. E. Siegman, "Hermite–gaussian functions of complex argument as optical-beam eigenfunctions," *JOSA*, vol. 63, no. 9, pp. 1093–1094, 1973. [Online]. Available: https://www.osapublishing.org/josa/abstract.cfm?uri=josa-63-9-1093

[24] M. A. Bandres and J. C. Gutiérrez-Vega, "Ince–gaussian beams," *Optics Letters*, vol. 29, no. 2, pp. 144–146, 2004. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-29-2-144

[25] J. C. Gutiérrez-Vega, M. D. Iturbe-Castillo, and S. Chávez-Cerda, "Alternative formulation for invariant optical fields: Mathieu beams," *Optics Letters*, vol. 25, no. 20, pp. 1493–1495, 2000. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-25-20-1493

[26] W. Paufler, B. Böning, and S. Fritzsche, "High harmonic generation with Laguerre-Gaussian beams," *Journal of Optics*, vol. 21, p. 094001, 2019.

[27] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde, "Transport-based model for turbulence-corrupted imagery," *Applied Optics*, vol. 57, no. 16, pp. 4524–4536, 2018.

[28] J. M. Nichols, A. T. Watnik, T. Doster, S. Park, A. Kanaev, L. Cattell, and G. K. Rohde, "An optimal transport model for imaging in atmospheric turbulence," 2017, arXiv:1705.01050.

[29] M. S.-E-Rabbi, X. Yin, A. H. M. Rubaiyat, S. K. S. Li, A. Aldroubi, J. M. Nichols, and G. K. Rohde, "Radon cumulative distribution transform subspace modeling for image classification," *CoRR*, vol. abs/2004.03669, 2020. [Online]. Available: https://arxiv.org/abs/2004.03669

[30] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde, "The cumulative distribution transform and linear pattern classification," *Applied and Computational Harmonic Analysis*, vol. 45, pp. 616–641, 2018.

[31] S. Kolouri, S. Park, and G. Rohde, "The radon cumulative distribution transform and its application to image classification," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 920–934, 2016.

[32] J. M. Nichols, M. N. Hutchinson, N. Menkart, G. A. Cranch, and G. K. Rohde, "Time delay estimation via Wasserstein distance minimization," *Signal Processing Letters*, vol. 26, no. 6, pp. 908–912, 2019.

[33] S. Kolouri, Y. Zou, and G. K. Rohde, "Sliced wasserstein kernels for probability distributions," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5258–5267.

[34] M. S.-E-Rabbi, X. Yin, A. H. M. Rubaiyat, S. K. S. Li, A. Aldroubi, J. M. Nichols, and G. K. Rohde, "Rcdt ns classifier," 2020, data retrieved from statista site, https://github.com/rohdelab/rcdt_ns_classifier/.

[35] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterington, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. [Online]. Available: http://proceedings.mlr.press/v9/glorot10a.html

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105. [Online]. Available: http://dl.acm.org/citation.cfm?id=2999134.2999257

[38] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv 1409.1556*, 09 2014.

[39] J. M. Nichols, T. H. Emerson, and G. K. Rohde, “A transport model for broadening of a linearly polarized, coherent beam due to inhomogeneities in a turbulent atmosphere,” *Journal of Modern Optics*, vol. 66, no. 8, pp. 835–849, 2019.

[40] V. I. Klyatskin and D. Gurarie, “Coherent phenomena in stochastic dynamical systems,” *Physics–Uspekhi*, vol. 42, no. 2, p. 165–19, 1999.

[41] F. Flandoli, M. Gubinelli, and E. Priola, “Well-posedness of the transport equation by stochastic perturbation,” *Inventiones Mathematicae*, vol. 180, pp. 1–53, 2010.

[42] P. Milanfar, “A model of the effect of image motion in the radon transform domain,” *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1276–1281, 1999.

CHAPTER 3

MACHINE LEARNING-BASED SIGNAL DEGRADATION MODELS FOR ATTENUATED UNDERWATER OPTICAL COMMUNICATION OAM BEAMS

Patrick L. Neary, [1][2] Abbie T. Watnik, [3] K. Peter Judd, [3] James R. Lindle, [4] and Nicholas S. Flann [1]

3.1 Abstract

We extend work from free-space, orbital angular momentum (OAM)-based communications to underwater optical communications (UWOC). Signal attenuation in UWOC is a problem that degrades classification performance. We develop and contrast several novel ways to create attenuation models and insert these models in the classification training pipeline. One model is built and trained using automatic differentiation and another uses the radon cumulative distribution transform (R-CDT). We show that including these attenuation models in classifier training significantly improves classification performance.

3.2 Introduction

Orbital Angular Momentum (OAM) in electromagnetic waves has created a stir in communications research ((1; 2)). OAM communications have great potential in free space or underwater ((3)) environments to increase communication bandwidth by multiplexing modes together. Early tests have shown Tera-bit/s potential data rates using four multiplexed modes ((4)). Several significant limiting factors in OAM communications exist, these include both turbulence ((5)) and signal attenuation ((6)).

While strides have been made in free space environments ((7; 8)), a number of issues remain in the underwater domain. (9) and (10) applied deep learning to the OAM domain

---

[1] Department of Computer Science, Utah State University, Old Main Hill, Logan, UT 84322, USA
[2] Space Dynamics Laboratory, 1695 N Research Park Way, North Logan, UT 84341
[3] Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA
[4] DCS Corporation, 6909 Metro Park Drive, Suite 500, Alexandria, VA 22310, USA

as a way to learn multiplexed OAM modes in free space. Both used a custom shallow CNN as well as AlexNet ((11)) for OAM mode classification. We extend the work presented in these papers to a signal attenuated in the underwater domain, and use the shallow CNN they presented as a baseline architecture. We are specifically interested in creating machine learning-based signal-noise models in the presence of attenuated signals, and using those models during training of a classifier, with the intent of improving classifier robustness.

The signal to noise ratio (SNR) between a laser source and an imager can change due to water turbidity. SNR in clear water may be initially high, but weather conditions may stir up particulates in the water and cause it to decrease. Additionally, the source and receiver may move to a location where SNR is lower. Any number of scenarios may cause a decrease in SNR. Consequently it's important to investigate system performance and robustness in these situations.

A potential weak point of OAM mode classification systems is how the classifier responds when presented with heavily attenuated, low SNR data. When training an OAM mode classifier, initial training sets are likely to be collected in environments different from those that the system will be actually operating in. Consequently, we are interested in finding ways to increase the robustness of the classification model for degrading factors such as turbulence or signal attenuation. In this work, we take a particular interest in signal attenuation.

As the term 'model' can mean refer to different things, we endeavor to clarify some terminology at the outset. The theoretical, physical model for attenuation is provided in Equation 3.10. We seek to create a machine learning-based model that learns how to take an image with a high SNR at its input and generate an image with a low SNR at its output. We refer to the machine learning-based model as '**S**NR **MA**chine lea**R**ning a**T**tenuation Model', or SMART Model for short. A SMART Model in this paper can be any machine learning-based model, such as a convolutional neural network (CNN). SMART models are discussed in Section 3.4.3.

We show that when a classifier is trained with high SNR images, it performs poorly on low SNR images. To improve performance, we insert a SMART Model in the training pipeline. This paper, therefore, explores the training of two distinct processes. First, we explore several methods for creating SMART Models as shown in Figure 3.1. Second, we create a standard CNN-based image classifier and try to improve its performance by using the SMART Model in the training pipeline, as depicted in Figure 3.2.

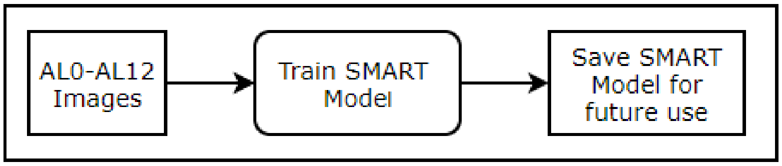


Fig. 3.1: High level training process for a SMART Model.

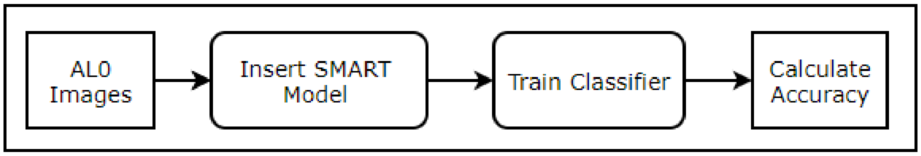


Fig. 3.2: High level training process for the classification model.

Concretely, contributions of this paper include the following points. We show that, starting from the baseline architectures of (10), when trained on a baseline data set and then presented with attenuated data, the classifier experiences serious degradation in performance. Where (10) worked with turbulent data in free-space communications, we're working with attenuated data in underwater communications. By creating, training, and inserting a SMART Model in the classifier training pipeline, we significantly improve the robustness of the classifier. We use several novel approaches to creating the SMART Models. The most compelling SMART model uses a Radon Cumulative Distribution Transform (R-CDT) ((10; 12; 13)), it achieves the best results while using only a single attenuated image from each class. As deep neural net models often require many samples, this approach could have significant implications in other machine learning applications as well.

3.3 Background and Prior Art

Orbital Angular Momentum

Orbital angular momentum in electromagnetic waves (EM) is a relatively new finding in physics. While known to exist at the atomic level, (14) identified and demonstrated its existence in 1992 using Laguerre-Gaussian (LG) beams. Electrons have long been known to exhibit spin (left or right rotation about the z-axis) and angular momentum, relating to the particular orbit of the electron. Allen and his team were able to show that monochromatic, EM waves also exhibit similar properties. EM waves are known to have spin or polarization. They were able to show that certain fields from Laguerre-Gaussian beams contain an angular component relative to the z-axis (direction of propagation), that is dependent on $exp(-i\ell\theta)$, where ℓ is called the topological charge. When $|\ell| > 0$ the wavefront becomes a helically shaped wavefront. The beam center, which usually has a radially Gaussian intensity profile, adopts a dark hole at the center (optical void) with an intensity pattern in a ring around the void. As the value of ℓ increases, the radius of the ring also increases.

Monochromatic light from an LG beam can be expressed with Equation 3.1:

$$\begin{aligned} U_{l,m}(\rho,\phi,z) = A_{l,m}[\frac{W_0}{W(z)}](\frac{\rho}{W(z)})^l L_m^l(\frac{2\rho^2}{W^2(z)})exp(-\frac{\rho^2}{W^2(z)}) \\ \times exp[-jkz - jk\frac{\rho^2}{2R(z)} \mp jl\phi + j(l+2m+1)\zeta(z)], \end{aligned} \tag{3.1}$$

The LG beam is defined in cylindrical coordinates (ρ, ϕ, z), where ρ is the radius, ϕ is the angle in the plane perpendicular to the z axis, and z is the axis of propagation. l and m are azimuthal and radial indices, $A_{l,m}$ is a constant, W_0 is the waist radius, $W(z)$ is the beam width, L_m^l represents generalized Laguerre polynomials, $R(z)$ is the radius of curvature for the wavefront at position z, and ζ is phase delay of the wavefront. The OAM dependency is given by $exp(-i\ell\phi)$, which controls whether the EM wave experiences a helical wavefront ($|\ell| > 0$) or a plane wave ($\ell = 0$).

A useful property of OAM, is that topical charge modes are orthogonal to each other ((15)). Consequently, modes of different topical charges can be multiplexed together, and data carried within the modes are completely orthogonal and easily demultiplexed, at least in ideal environmental conditions.

Several methods exist for imparting OAM modes to Gaussian laser beams ((16; 17; 18)). Optical vortex phase plates are an optic composed of a helical surface that is specially designed for specific wavelengths. Each phase plate is designed to impart a specific mode to the Gaussian beam.

Optimal Transport

Mathematics of optimal transport, or Earth Mover's distance, can be applied to intensity information from signals, among other application domains ((19)). (20) formulated a link between atmospheric propagation of EM waves to optimal transport. EM radiation can be modeled with Maxwell's equations for isotropic materials. Given Maxwell's equations (3.2-3.5), if we take the curl of Equation 3.2 and substitute in Equations 3.3-3.5 we get Equation 3.6.

$$\bigtriangledown \times \mathbf{E}(x) = i\omega\mu_0\mathbf{H}(x) \tag{3.2}$$

$$\bigtriangledown \times \mathbf{H}(x) = -i\omega\epsilon_0\epsilon(x)\mathbf{E}(x) \tag{3.3}$$

$$\mu_0 \bigtriangledown \cdot\mathbf{H}(x) = 0 \tag{3.4}$$

$$\mu_0 \bigtriangledown \cdot(\epsilon(x)\mathbf{E}(x)) = 0 \tag{3.5}$$

$$\bigtriangledown^2\mathbf{E}(x) + \bigtriangledown(\mathbf{E}(x) \cdot \frac{\bigtriangledown\epsilon(x)}{\epsilon(x)}) + k_0^2\epsilon(x)\mathbf{E}(x) = 0. \tag{3.6}$$

Substituting the electric field with the scalar electric field ($\mathbf{E}(x) = \Psi(\vec{x}, z)e^{ik_0z}$) in Equation 3.6 yields an equation similar to the Schrodinger equation. The Madelung transformation

(21) can then be used on it to derive the following equations

$$\frac{\partial \rho(\vec{x},z)}{\partial z} + \bigtriangledown_X \cdot [\rho(\vec{x},z)v(\vec{x},z)] = 0 \tag{3.7}$$

$$\frac{\partial \rho(\vec{x},z)}{\partial z} + [v(\vec{x},z) \cdot \bigtriangledown_X]v(\vec{x},z) = 2 \bigtriangledown x\gamma[\eta(\vec{x},z)] \tag{3.8}$$

$$\frac{\partial \phi(\vec{x},z)}{\partial z} + \frac{1}{2}[\bigtriangledown_X \phi(\vec{x},z)]^2 = 2\gamma[\eta(\vec{x},z)]. \tag{3.9}$$

Equations 3.7, 3.8, and 3.9 describe electric field propagation through the atmosphere in terms of intensity and phase. The format of Equations 3.7-3.9 allow electric fields to be modeled and analyzed using optimal transport theory ((22)).

(10) showed that, given the formulation of electric fields in a form compatible with optimal transport, one can create a transform between a clear and corrupted image. An efficient method for computing the transform is called the Radon Cumulative Distribution Transform ((23), (5)). The R-CDT is used in this work to provide transforms between high and low SNR images.

OAM Communications

(10) presented free space OAM mode classification in the presence of turbulence. They trained classifiers using OAM images and compared them to classifiers trained using patterns from the R-CDT. We use their work as a launching point for exploring the UWOC domain. Specifically, they used a custom neural network and showed very good classification results with it. That same architecture will be used here. They used the R-CDT patterns for classification, but in this paper, the R-CDT will be used to map intensity distributions from low to high attenuation levels.

While neural networks have been able to achieve impressive results in many different applications, one concern with using them is that their predictions may be inconsistent with physical realities. This has been one of the drivers behind the idea of physics-guided training in neural networks. (24), for example, demonstrated a physics-guided neural network that was used for modeling lake temperatures. During training, it was shown that the neural network could predict values inconsistent with reality. By including physics models in the training loop, they were able to speed up and improve the model's results by using physics models to impose limits on realistic outputs.

We adapt this principle to training an OAM classification model. As it has been shown that optimal transport is able to capture the underlying physics in a system ((20)), we can use it to learn a physics-based SMART model for attenuation. An advantage of using the R-CDT is that it is able to capture the transform using a minimal number of samples. Once the SMART model is learned, it can be placed in the training loop. The SMART model can be used to modify a high SNR data set and transform some of the images to their low SNR counterparts. We show that applying these principles greatly improves classification performance.

The rest of this paper will proceed as follows. In Section 3.4 we discuss the hardware configuration used to generate OAM data and details, attenuation model creation, and classification network architecture with training parameters. Section 3.5 presents results from training and inference. Finally, Sections 3.6 and 3.7 conclude with discussions of the results and conclusion.

3.4 Experiment Setup

In this Section we discuss the hardware configuration used to generate the OAM images, image set composition, SMART model architectures and training details, OAM mode classifier architecture and training details, and final test configuration.

3.4.1 Hardware Configuration

A 1.2 meter water tank was used in the configuration for image acquisition. A Laguerre-Gauss beam is generated by a Q-switched diode pumped solid state laser operating at 532 nm. The laser and a camera were are synchronized at 1 KHz. The laser beam is split into four coherent beams and is passed through vortex phase plates with different charges, where an OAM phase is imparted to each beam. Subsequently, the beams are recombined using beamsplitters, the beam passes through the water tank, and are incident on the camera. The OAM modes used for this configuration are [1, 4, -6, -8].

A Phototron FASTCAM SA-Z camera was externally triggered to be in sync with the laser. Several unique artifacts seen in the images are associated with camera dynamics. In Figure 3.4, one may notice vertical stripes in the images. They are very apparent in the higher attenuation length images. Additionally, the image intensity patterns appear to be pixelated, as opposed to being continuous. Both of these characteristics are byproducts of camera dynamics.

Since OAM modes are orthogonal, the modes can be mixed and matched to produce different light patterns. The beams for each phase plate can be blocked, so with 4 OAM modes there are 2^4 possible combinations. For this effort, 15 patterns are used and one class consists of a background reading, which is an image capture on the camera when the laser beam is completely blocked. One thousand images were collected for each class (or OAM pattern), for a total of 16,000 images per data collection. Figure 3.3 shows intensity patterns, without attenuation, for each of the OAM mode combinations used in this paper.

3.4.2 Image Set Composition

Five micron plastic beads were added to the water to attenuate the signal by different amounts. The beads are designed to absorb or scatter the light, and the amount of absorption and scattering is proportional to the number of beads present. A total for 4 sets of data collects were made for attenuation lengths 0, 4, 8, and 12 (AL0, AL4, AL8, AL12), where AL0 was collected without any attenuation beads present. Signal attenuation lengths are described by Equation 3.10, where i is the attenuated signal, i_0 is the original signal strength, and AL is the level of attenuation. In this case AL values are set to 0, 4, 8, and 12.

$$i = i_0 e^{-AL} \tag{3.10}$$

Figure 3.4 shows an example of six classes at each of the attenuation lengths. In OAM pattern 0001, the AL0 image shows a distinct pattern with a high SNR. As the attenuation length increases, the noise floor of the camera moves up and the SNR drops. At AL12 we

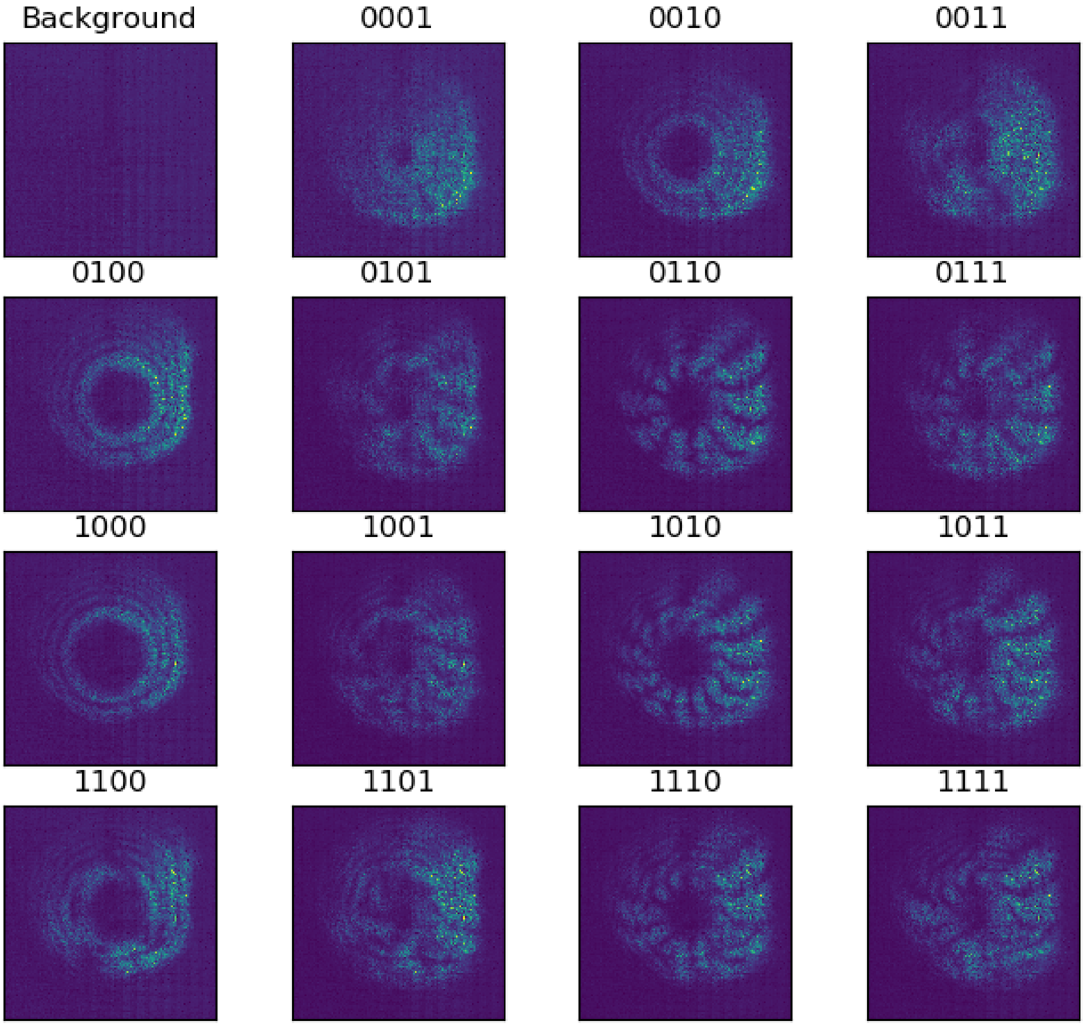


Fig. 3.3: Examples of intensity patterns for each OAM mode combination.

see the noise overcoming the lower intensity portions of the OAM pattern. For a standard convolutional neural network trained to look for patterns, it is easy to see why there would be significant degradation of performance when presented with images with low SNR.

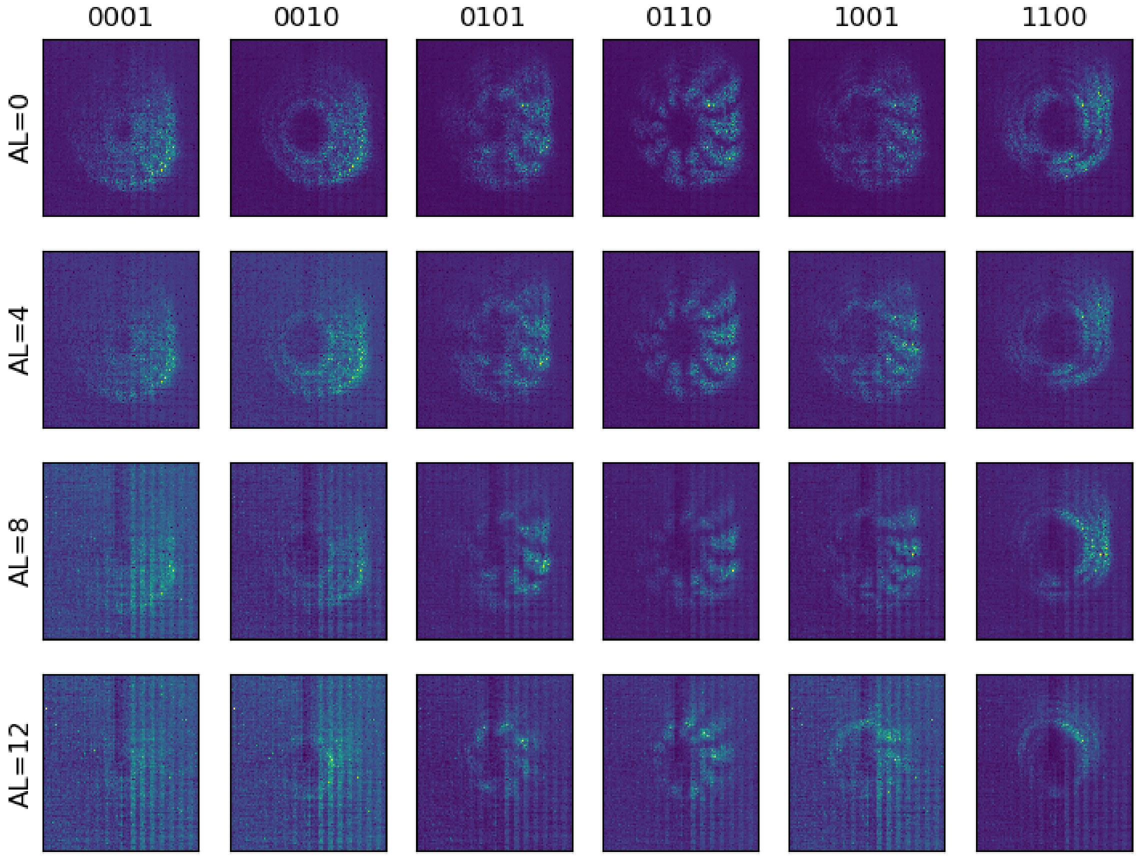


Fig. 3.4: Examples of attenuation in underwater environment. Each column contains patterns for one OAM mode and each row represents a specific level of attenuation. The first row contains attenuation level 0. The subsequent rows show progressively higher levels of attenuation. Note that these images are scaled to the brightest pixel in each individual image to make the OAM mode patterns easier to see.

In generating results for this paper, the data sets were divided up in the following way. Each attenuation set was split into 70%/15%/15% for training, testing, and validation, respectively. The attenuation and classification models were trained using only the training and validation sets. Test sets were used only once, at the end to produce final results. Images from the camera were originally 1024x1024 grayscale images. The borders were cropped to remove pixels without intensity information and were downsampled to a final dimension of 128x128.

As we are interested in seeing how well the classifier performs with unseen data, the OAM mode classifier is trained using only the AL0 data set. Once the classifier has been trained, its accuracy is evaluated with all attenuated image test sets. Concretely, the AL0 test set is first passed through the classifier and its performance metrics are recorded. AL4 is then passed through the classifier and its performance metrics are recorded. AL8, and AL12 are sequentially processed in the same manner. This way we are able to see how well the classifier performs with each group and can easily see the performance trends as the attenuation level increases. While the sets were processed serially, they could have been processed in parallel with the right hardware.

3.4.3 SMART Model Details

The purpose of the SMART attenuation model is to learn how to mimic the SNR characteristics of the environment and hardware. Once an algorithm learns that mapping, the SMART model can be placed in the classifier training loop. In this section, we discuss details of the SMART models used to learn signal attenuation performed by the environment. Two different SMART models are discussed, a convolution based model and an R-CDT based model.

CNN-Based SMART Model

For the CNN SMART model, we combine convolutions and feature maps in a non-standard way to produce images that resemble their attenuated counterparts. Figure 3.5 shows the architecture for the CNN model. Model construction is based on observations that AL4-AL12 images have a low SNR and the camera introduces pixelated characteristics in the intensity patterns and background. The first convolution and max layers act on the original image and then the original image is subtracted from the result. There are several benefits to this combination. One, the effective SNR between background and signal is reduced. Additionally, as previously mentioned, the camera being used introduces a pixelated characteristic in the intensity patterns. Convolutions can potentially smooth out an image, so subtracting off the original image can reintroduce some of those pixelated characteristics. This process is repeated in several stages as displayed in Figure 3.5.

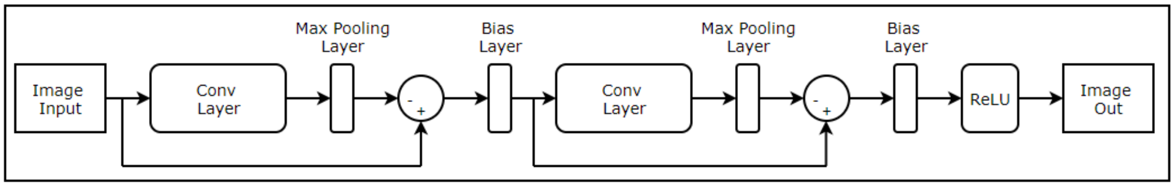


Fig. 3.5: CNN model architecture.

In training the CNN-based SMART attenuation model, all AL training data sets (AL0, AL4, AL8, and AL12) are used. The four sets contain a total of 54,400 images. During training, AL0 images are randomly selected from the training set. Once the source images have been collected, attenuated images are drawn from matching classes and paired with the source images. Loss terms are calculated using the mean-squared-error (MSE) between the model output and images from the attenuated set. Model weights are updated using automatic differentiation (25) with the calculated loss. In this way, the CNN model learns how to map from one attenuation level to another. This process is shown in Figure 3.6.

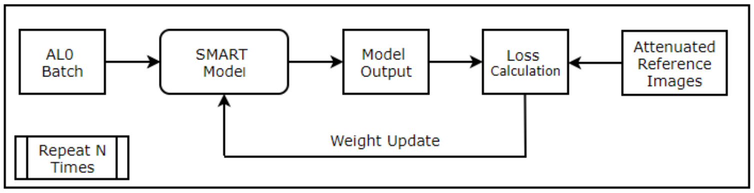


Fig. 3.6: Training loop for the CNN based model.

Automatic differentiation is used in backpropagation, but can be generally applied to any equation. In this case, the CNN-based model is trained using automatic differentiation. Once an image is generated by the model, a difference between the two images is calculated using the MSE from pixel differences in the images. The loss calculated from the MSE is used to generate gradients, which are then applied to the weights in the model. Training of the CNN model takes place over 100 epochs with a learning rate set to $1e^{-4}$.

Figure 3.7 contains an example of the output from this architecture. The image on the left is the AL0 input, the image in the middle is the AL12 reference, and the image on the right is the model output.

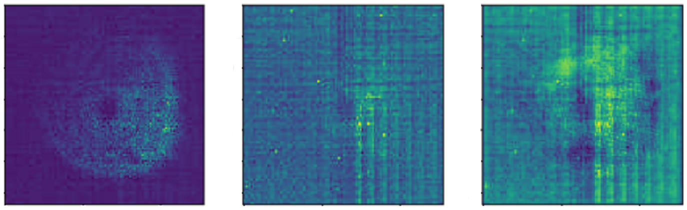


Fig. 3.7: CNN AL0 input (left), AL12 reference (center), and CNN forward map output (right).

R-CDT-Based SMART Models

As previously discussed, the R-CDT, learns a mapping from one mass distribution to another. In this case, the initial intensity distribution is provided by image classes in the AL0 training set. The 'final' intensity patterns are provided by corresponding classes in the attenuated data sets.

Once the R-CDT has been computed, it creates a transform that enables a translation of the intensity 'mass' from one OAM AL level to another. The R-CDT enables either forward or backward transforms, so as an example, one might have an image without attenuation, apply the forward transform, and get an image with intensity mass distribution of a high attenuation level.

In this work, two R-CDT SMART models were created. The first model is a mapping from AL0 images to AL12 images, so it uses 32 total images to form the model (16 from AL0 and 16 from AL12). With this model, we look to learn whether AL4 and AL8 also see improved gains in accuracy or just AL12. The second model includes mappings from AL0 to AL4, AL8, and AL12. This model is created by using a single source image from each class in AL0 and mapping them to corresponding images from the classes in the other attenuated sets. This model uses a total of 64 images to create the mappings. This is in contrast to the 54,400 images used to train the CNN model.

Figure 3.8 shows an example of the R-CDT input, reference, and output images. The left image is from the AL0 set and is given to the R-CDT forward map operation. The middle image is an example of the same class from the AL12 data set. The right image is the output from the forward operation of the R-CDT model.

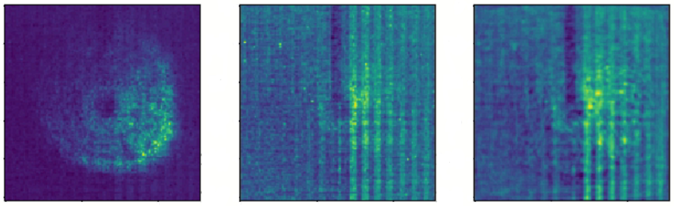


Fig. 3.8: R-CDT AL0 input (left), AL12 reference (center), and R-CDT forward map output (right).

3.4.4 Classification Model Details

The training classifier is a shallow convolutional network patterned after the one used in (10). Using this same architecture provides a baseline for comparison to previous work. The shallow convolutional network is composed of six basic layers. It has a convolution layer with 16, 11x11 kernels, a convolution layer with 32, 3x3 convolution kernels, a max pooling layer, a ReLU, a flattening operation, and a dense layer with an output size of 16 (one for each class). The optimizer is Adam with an initial learning rate of $1e^{-3}$. The mini-batch size is 32 and training takes place over 1050 iterations.

The classifier training process is shown in Figure 3.9. While training the classifier, training batches from AL0 are acquired. Images within the batch are then randomly selected to be transformed with the SMART model. Random selection helps broaden the variety of images the classifier sees over multiple epochs of training. The batch is then passed on to the classification model for training.

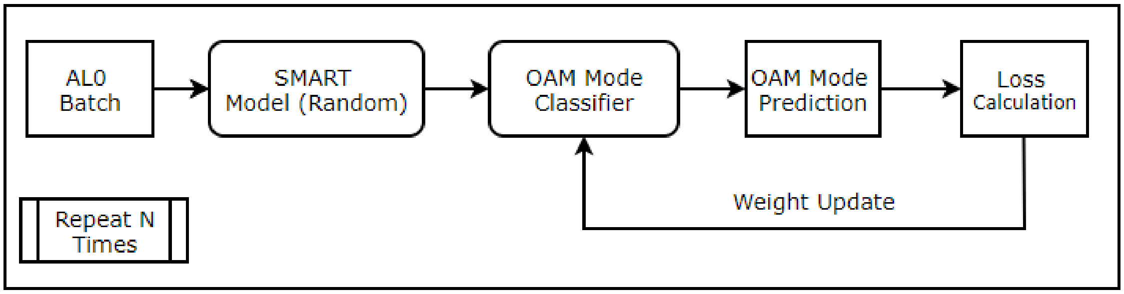


Fig. 3.9: Training loop for the classification model.

Figure 3.10 shows the process for using the attenuation model in the training pipeline. At the top, it shows the different attenuation length sets. The left box indicates a test set that will be fed through the classifier. The data goes through the classifier, generates a set of predictions, and then the results are accumulated and a confusion matrix is created based on those results. Test sets are passed, one at a time, through the classifier and results are collected for each set.

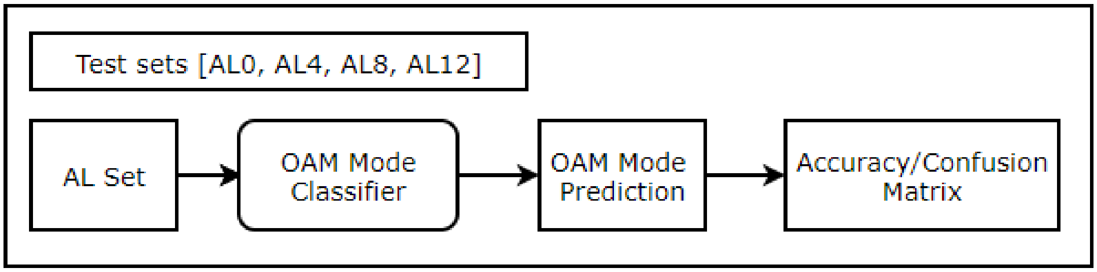


Fig. 3.10: Test loop for the classification model.

3.5 Results

The CNN and R-CDT-based SMART models were trained on images from the AL0, AL4, AL8, and AL12 training sets. After training the CNN-based SMART model, it was placed in the classification training pipeline and randomly enabled to apply transformations to the high SNR images. The transformed images were then passed on as inputs to the classifier for training. Likewise, once the R-CDT-based SMART attenuation model was created, it was placed in a new classification training model pipeline and randomly enabled to apply transformations on the high SNR images.

Table 3.1: **Classifier comparison**

Classifier	AL0	AL4	AL8	AL12
No model	100	75.9	23.1	6.1
CNN model	99.6	82.6	66.5	55.5
R-CDT model 1	99.9	86.8	65.3	73.5
R-CDT model 2	99.9	97.3	82.6	74.2

Table 3.1 shows results from inference using the different SMART models in the classification architecture.

The first entry shows results for the classifier that is trained without a SMART model in the training pipeline. This provides the baseline results to see how much improvement is made by adding a SMART model into the training baseline. As expected, results for AL0 are very good. AL4 sees a drop to 75.9%. AL8 and AL12 accuracies become unusable at 23.1% and 6.1%.

The second entry in Table 3.1 displays results for the CNN-based SMART model. AL0 experiences a small drop in accuracy, however AL4 jumps to 82.6% while AL8 and AL12 also experience significant jumps in accuracy to 66.5% and 55.5%.

Row three in the table shows results for the first R-CDT SMART model. This model contains the single mapping from AL0 to AL12. AL0 shows improved performance over the CNN model. It also shows better results for AL4 at 86.8%, slightly lower results for AL8 at 65.3%, and a significant improvement in AL12 at 73.5%.

Row four contains results from the second R-CDT SMART model, with mappings from AL0 to AL4, AL8, and AL12. This model shows the best results with AL0 at 99.9%, AL4 at 97.3%, AL8 at 82.6%, and AL12 at 74.2%.

Table 3.2 shows a confusion matrix generated from the classifier using second R-CDT SMART model. This table provides some insight into where the majority of the problems exist for the AL12 data set. AL12 was selected because it performed the worst and may provide the most insight into ways to improve. In this table, ‘0’ corresponds to the ‘Background’

Table 3.2: Confusion matrix for AL12 classification with R-CDT model 2.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	246	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	237	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	232	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	64	0	169	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	219	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	227	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	22	15	1	0	36	0	8	0	0	0	158	0
7	0	0	0	0	1	187	0	44	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	0	212	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	20	208	0	0	0	0	0	21
10	0	0	0	0	0	0	0	0	4	0	237	0	0	0	1	0
11	0	0	0	0	0	0	0	0	20	29	87	78	2	0	11	3
12	0	0	0	0	7	0	0	0	49	0	0	0	104	0	77	0
13	0	0	0	0	35	1	0	0	10	46	0	0	31	93	8	23
14	0	0	0	0	0	0	0	0	0	0	4	0	3	0	222	0
15	0	0	0	0	8	0	0	0	23	33	2	3	0	7	151	4

image from Figure 3.3, ‘1’ corresponds to ‘0001’, and so on. From the table we can see that the images 6 (0110), 7 (0111), 11 (1011), 13 (1101), and 15 (1111) performed poorly.

3.6 Discussion

As shown in Table 3.1, the base classifier was unable to perform at the AL8 and AL12 attenuation levels, while having moderate results at the AL4 attenuation. Once the SMART models were introduced, accuracy for all attenuation levels jumped and AL12 experienced the greatest improvement by jumping from 6.1% to 74.2%.

It’s interesting to note the significant improvement in performance of the R-CDT-based SMART model over the CNN SMART model. While the CNN-based model used all available training data, the R-CDT model used only one image from each of the classes in the AL12 based set. One of the difficulties that is experienced in machine learning applications is the collection and labeling of training data sets. Indeed, in many applications where groups desire to apply machine learning, the collection and labeling of large data sets for training represent a significant hurdle. The fact that the R-CDT SMART model was able to generate such a significant improvement in performance with so few examples is a powerful result. This may have some potential implications in the areas of one-shot and few-shot learning.

From the confusion matrix in Table 3.2, we observe a number of classes with particularly poor performance. The poor performance is very likely due to attenuated signals whose patterns are very similar to each other. It shows that patterns 6, 7, 11, 12, 13, and 15 have the majority of the problems. Looking more closely at 6 (0110), for example, shows that most of the misclassifications are directed at 14 (1110). Inspecting the two patterns visually, it's apparent that the two closely resemble each other. In future efforts, this information can be used to help select OAM modes and patterns with characteristics that differ more under high attenuation environments.

3.7 Conclusion

From the outcomes presented in Section 3.5, we observed a significant improvement in OAM classification results under heavy signal attenuation. Without an attenuation model, the baseline classifier experienced a 6.1% accuracy level for the AL12 data set. With the best SMART model in the training loop, the accuracy level was raised to 74.2%.

We were able to see significant improvement in accuracy for attenuation levels the classifier had not previously seen. This was accomplished by inserting a SMART model into the classifier training pipeline. The best performing model was based on the R-CDT, which has strong fundamental ties to the underlying physics of the optimal transport of photons.

A significant characteristic of the R-CDT based models is the low number of samples required. The first model required only 32 images to perform the mapping between AL0 OAM patterns and AL12 OAM patterns. The second model required only a few more at 64 images to map from AL0 to AL4, AL8, and AL12. The ability of the R-CDT to create such a robust SMART model with so few images may have significant implications in machine learning, especially for applications dependent on one-shot or few-shot learning. That is an area that warrants future research.

The confusion matrix for the second R-CDT model is shown in Table 3.2. It provides future direction as to which patterns to consider changing to provide greater pattern discrimination in high attenuation environments.

REFERENCES

[1] R. Chen, H. Zhou, M. Moretti, X. Wang, and J. Li, "Orbital angular momentum waves: Generation, detection and emerging applications," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[2] A. E. Willner, Z. Zhao, Y. Ren, L. Li, G. Xie, H. Song, C. Liu, R. Zhang, C. Bao, and K. Pang, "Underwater optical communications using orbital angular momentum-based spatial division multiplexing," *Optics Communications*, vol. 408, pp. 21–25, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401817306818

[3] J. Baghdady, K. Miller, K. Morgan, M. Byrd, S. Osler, R. Ragusa, W. Li, B. M. Cochenour, and E. G. Johnson, "Multi-gigabit/s underwater optical communication link using orbital angular momentum multiplexing," *Optics Express*, vol. 24, no. 9, pp. 9794–9805, May 2016. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-9-9794

[4] H. Huang, G. Xie, Y. Yan, N. Ahmed, Y. Ren, Y. Yue, D. Rogawski, M. J. Willner, B. I. Erkmen, K. M. Birnbaum, S. J. Dolinar, M. P. J. Lavery, M. J. Padgett, M. Tur, and A. E. Willner, "100 tbit/s free-space data link enabled by three-dimensional multiplexing of orbital angular momentum, polarization, and wavelength," *Optics Letters*, vol. 39, no. 2, pp. 197–200, 2014. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-39-2-197

[5] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde, "Transport-based model for turbulence-corrupted imagery," *Applied Optics*, vol. 57, no. 16, pp. 4524–4536, 2018. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-57-16-4524

[6] B. Cochenour, K. Dunn, A. Laux, and L. Mullen, "Experimental measurements of the magnitude and phase response of high-frequency modulated light underwater,"

Applied Optics, vol. 56, no. 14, pp. 4019–4024, 2017. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-56-14-4019

[7] T. Doster and A. T. Watnik, “Measuring multiplexed OAM modes with convolutional neural networks,” in *Lasers Congress 2016 (ASSL, LSC, LAC) (2016), paper LTh3B.2*. Optical Society of America, 2016, p. LTh3B.2. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=LSC-2016-LTh3B.2

[8] S. Lohani, E. M. Knutson, M. O’Donnell, S. D. Huver, and R. T. Glasser, “On the use of deep neural networks in optical communications,” *arXiv:1806.06663 [physics, physics:quant-ph]*, 2018. [Online]. Available: http://arxiv.org/abs/1806.06663

[9] T. Doster and A. T. Watnik, “Machine learning approach to OAM beam demultiplexing via convolutional neural networks,” *Applied Optics*, vol. 56, no. 12, pp. 3386–3396, 2017.

[10] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, “Demultiplexing vortex modes in optical communications using transport-based pattern recognition,” *Optics Express*, vol. 26, no. 4, pp. 4004–4022, 2018.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[12] S. Kolouri, S. R. Park, and G. K. Rohde, “The radon cumulative distribution transform and its application to image classification,” *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 920–934, 2016. [Online]. Available: http://arxiv.org/abs/1511.03206

[13] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde, “The cumulative distribution transform and linear pattern classification,” *Applied and Computational Harmonic Analysis*, vol. 45, no. 3, pp. 616–641, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520317300076

[14] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, "Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes," *Physical Review A*, vol. 45, no. 11, pp. 8185–8189, 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.45.8185

[15] W. Cheng, W. Zhang, H. Jing, S. Gao, and H. Zhang, "Orbital angular momentum for wireless communications," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 100–107, 2019.

[16] N. R. Heckenberg, R. McDuff, C. P. Smith, and A. G. White, "Generation of optical phase singularities by computer-generated holograms," *Optics Letters*, vol. 17, no. 3, pp. 221–223, 1992. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-17-3-221

[17] M. W. Beijersbergen, L. Allen, H. E. L. O. van der Veen, and J. P. Woerdman, "Astigmatic laser mode converters and transfer of orbital angular momentum," *Optics Communications*, vol. 96, no. 1, pp. 123–132. [Online]. Available: http://www.sciencedirect.com/science/article/pii/003040189390535D

[18] M. W. Beijersbergen, R. P. C. Coerwinkel, M. Kristensen, and J. P. Woerdman, "Helical-wavefront laser beams produced with a spiral phaseplate," *Optics Communications*, vol. 112, no. 5, pp. 321–327. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0030401894906386

[19] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 43–59, 2017.

[20] J. M. Nichols, A. T. Watnik, T. Doster, S. Park, A. Kanaev, L. Cattell, and G. K. Rohde, "An optimal transport model for imaging in atmospheric turbulence," 2017, arXiv:1705.01050.

[21] P. Vadasz, "Rendering the navier-stokes equations for a compressible fluid into the schrödinger equation for quantum mechanics," *Fluids*, vol. 1, no. 2,

p. 18, 2016. [Online]. Available: https://nau.pure.elsevier.com/en/publications/rendering-the-navier-stokes-equations-for-a-compressible-fluid-in

[22] J.-G. Liu, R. L. Pego, and D. Slepčev, "Least action principles for incompressible flows and geodesics between shapes," 2016, arXiv:1604.03387.

[23] S. Kolouri, S. R. Park, and G. K. Rohde, "The radon cumulative distribution transform and its application to image classification," *CoRR*, vol. abs/1511.03206, 2015. [Online]. Available: http://arxiv.org/abs/1511.03206

[24] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (pgnn): An application in lake temperature modeling," 2017, arXiv:1710.11431.

[25] B. Günes, P. A, R. Andreyevich, and S. Mark, "Automatic differentiation in machine learning," *The Journal of Machine Learning Research*, 2017. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3122009.3242010

CHAPTER 4

OPTIMAL ARCHITECTURES FOR OAM BASED COMMUNICATIONS IN ATTENUATED UNDERWATER AND TURBULENT FREE-SPACE COMMUNICATIONS

Patrick L. Neary, [1][2] Abbie T. Watnik, [3] K. Peter Judd, [3] James R. Lindle, [4] and Nicholas S. Flann [1]

4.1 Abstract

Turbulence and attenuation are signal degrading factors that can severely hinder free-space and underwater OAM optical pattern demultiplexing. A variety of state-of-the-art convolutional neural network architectures are explored to identify which, if any, provide optimal performance under these non-ideal environmental conditions. Hyperparameter searches are performed on the architectures to ensure that near-ideal settings are used for training. Architectures are compared in various scenarios and the best performing, with their settings, are provided. We show that from the current state-of-the-art architectures, DenseNet outperforms all others when memory is not a constraint. When memory footprint is a factor, ShuffleNet is shown to performed the best.

4.2 Introduction

In 2014, Krenn et al. explored the use of machine learning (ML) to demultiplex OAM beam patterns for free-space optical communications (1). Since then, ML techniques have been applied in a variety of ways to improve demultiplexing accuracy in free-space turbulent conditions (2; 3; 4).

---

[1] Department of Computer Science, Utah State University, Old Main Hill, Logan, UT 84322, USA
[2] Space Dynamics Laboratory, 1695 N Research Park Way, North Logan, UT 84341
[3] Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA
[4] DCS Corporation, 6909 Metro Park Drive, Suite 500, Alexandria, VA 22310, USA

In OAM communications, turbulence and attenuation can cause significant degradation of signal integrity and lowering of the signal-to-noise ratio (SNR) (5; 6). These disturbances can displace spatial patterns, cause crosstalk, or scatter the signals such that only a portion of the original intensity distribution makes it to the receiver.

One of the unresolved questions from Ref. (2) is with regards to which, if any, of the state-of-the-art convolutional neural network (CNN) architectures performs best for OAM pattern demultiplexing in signal degrading environments. This paper explores turbulent free-space and attenuated underwater OAM optical communications with the state-of-the-art deep convolutional neural networks to answer this question.

Several data sets under varying environmental conditions are used for this effort. In free-space, three sets of data are collected at different turbulence levels. In water, four sets of data are collected at various attenuation levels. All tests are performed on specific combinations of these data sets.

Contributions of this paper include a comparison of recent, state-of-the-art CNN architecture in both turbulent free-space and attenuated underwater OAM communications. Baseline performance, inter-set performance, and parameter count are analyzed. At the end of the analysis, the best performing architectures, along with their parameters, are provided.

4.3 Background and Prior Art

In the following sections OAM communications, hyperparameter tuning, and an overview of some of the current state-of-the-art CNNs are covered.

4.3.1 Orbital Angular Momentum

Orbital angular momentum (OAM) in electric fields was discovered by Allen et al. (7). They found that under certain conditions, the Laguerre-Gauss beam could transition from a standard plane wave propagation to a helical path. Consequently, the Gaussian-shaped distribution frequently exhibited by lasers becomes a doughnut shaped pattern when an OAM mode is adopted. The OAM azimuthal dependency is expressed by $exp(i\ell\phi)$, where ℓ is the topological charge or mode number. When $\ell = 0$, the wavefront is a plane. When $|\ell| > 0$, the wavefront travels in a helical path, where the direction of rotation about the z-axis is controlled by the sign on ℓ. The radial distance from the z-axis to the helix is controlled by the mode number. The larger the mode number, the greater the radius.

A significant property exhibited by OAM modes is that they are orthogonal to each other (7). Consequently, multiple OAM beams with different modes can be multiplexed together and be completely recovered at the receiver. Leveraging this property allowed Want et al. to achieve terabit data rates in ideal conditions (8). While promising, communications in non-ideal conditions with turbulence and attenuation present hurdles in actually achieving these rates.

Various of approaches have been used to detect modes at the receiver, including conjugate mode sorting (9; 10), Doppler effect measurements (11), dove prism interferometers (12), optical transformers (13), and spiral fringe counting (14). In 2017 Doster and Watnik applied ML to the problem and found significant improvements in demultiplexing accuracy and simplification of hardware setup over existing approaches (2).

While Ref. (2) provided a proof of concept in applying CNNs to determining OAM modes, they left an exploration of the best CNN for future work. That investigation is completed here in addition to addressing other questions.

4.3.2 State-of-the-Art CNNs

CNNs are great for image-based applications because their convolution kernels are able to learn and differentiate shapes, colors, hues, etc. found in the training images. The unique characteristics associated with each class are learned during training and then used later for identifying classes during inference.

The deep learning revolution was accelerated with the groundbreaking results achieved by the LeNet architecture developed by LeCun (15). The AlexNet (16) architecture later smashed previous records on the ImageNet (17) data set by combining convolution layers, fully connected layers, rectified linear units (ReLU), and dropout layers. ImageNet is a benchmark set of images containing 1,000 different classes and is frequently used to compare architecture performance.

Since AlexNet, a number of significant improvements have been made in CNN architectures, layers, and optimizers. Adam, for example, is an optimizer that provides adaptive learning rates (18). As training progresses, learning rates are automatically adjusted up or down so weights do not train too slowly or diverge, because learning rates are too high.

ResNet created another revolution in the CNN architecture field by introducing the concept of special processing blocks surrounded by identity connections (19). ResNet is able to address the problem of diminishing gradients by allowing vital image information to be made available, through the identity connections, to deeper layers of the network. ResNet is able to bypass the accuracy inflection point of the VGG architecture, where adding layers caused decreases in performance (20). ResNet is able to continue improving performance by stacking much deeper layers, beyond which VGG performance degraded.

ResNeXt is an extension of the ResNet architecture (21). They postulated that gains could be made through widening the architecture. They introduced the idea of cardinality, where N branches were introduced and each branch contained a small number of kernels.

DenseNet provides a complimentary approach to ResNet (22). Rather than using identity connections, DenseNet is able to feed forward feature maps from each processing block (which is composed of convolution layers and other operations). Each processing block is provided feature maps, at its input, from all previous processing blocks.

SqueezeNet is an architecture designed for small memory footprint applications (23). It is composed of 'Fire modules' which contain squeeze (1x1 convolutions) and expand (1x1 and 3x3 convolution) layers. It was found to perform comparable to AlexNet on the ImageNet set, but with 50x fewer parameters.

MobileNet is a series of streamlined architectures that use depth-wise separable convolutions (24). The series was designed to be lightweight so as to be appropriate for applications such as self-driving cars, robotics, etc.

ShuffleNet is an architecture designed to be memory efficient, so as to be deployed in robotics and mobile devices (25). While its accuracies are not competitive with architectures like ResNet, when trained and tested against ImageNet, it has a much lighter memory footprint and can fit in small devices where the larger architectures will not. It was shown to have better accuracies on the ImageNet set than MobileNet.

SqueezeNet, MobileNet, and ShuffleNet are small, efficient architectures where AlexNet, DenseNet, ResNet, ResNeXt, and WideResNet are large but produce better results on the ImageNet benchmark data set. These architectures will be used and contrasted with each other in this work to determine which provides the best performance in the OAM communication domain. Ref. (26) provides an interesting insight that some of the state-of-the-art CNNs struggle in their ability to generalize and are in fact 'brittle'. This warrants a careful consideration in applying the state-of-the-art to new domains. The best architecture in ImageNet classification does not necessarily mean optimal performance in another domain. These architectures are trained and analyzed to show which perform best in OAM free-space and underwater communications.

4.3.3 Hyperparameter Tuning

'Hyperparameters' are parameters that are set before training begins. Examples of these parameters include learning rates, batch size, number of training epochs, optimizer, methods for weight initialization, etc.

Learning rates control the magnitude of updates to weights during training. When learning rates are too large, training can diverge. If weights are too small, it may take a very long time to converge to a solution.

An epoch is a cycle of training where all available data has been used once to train the network. A data set is often broken into batches and trained one batch at a time. The batch size can influence how well the architecture trains. The number of epochs also influences the overall performance of the network. If the network trains for too many epochs it can over train and will not generalize well. If trained too little, then the network will not learn the unique characteristics of the information presented to it. Either case can result in poor performance.

Adam is currently among the most popular optimizers (18). Ref. (27) found that Adam did not always perform better than other optimizers. In light of this, several additional optimizers were selected to include in the parameter search. A comparison of various optimizers was done by Ruder (28). From their list, the following optimizers were selected for comparison: Adam, AdaMax (18), Nadam (29), and RMSProp (30).

In order to perform a fair comparison between the selected CNN architectures, a hyperparameter search can be used to make sure architectures have been suitably configured. Selecting a good learning rate is considered to be one of the most important hyperparameters to tune (31). For this research, optimizers and learning rates are evaluated.

Approaches to hyperparameter tuning are an ongoing area of research. The most basic approach is a grid search, where a range of values is selected for each hyperparameter, the values are changed one at a time, and all combinations are exhaustively evaluated. This approach is computationally expensive. A more effective approach is a random search (32). It is able to find better configurations in less time compared to grid search.

Hyperopt is a parameter tuning approach that allows searching across multiple hyperparameters in an efficient way (33). Ref. (33) shows that Hyperopt provides an order of magnitude speed-up over Bayesian parameter tuning methods. For this research, the Hyperopt algorithm was used from the Tune package for the learning rate search (34).

4.4 Experiment Setup

Turbulent free-space image sets and attenuated underwater sets of data are used in this study. Free-space turbulent data is collected according to the setup described in Ref. (2). This imagery is collected using a 635-nm 5-mW laser, a Dalsa GigE camera, several Forth Dimension Displays binary phase ferroelectric SLMs, and standard optical tools such as mirrors, pinhole filters, and diffraction order filters. The SLMs are programmed with a binary phase hologram. Simulated turbulence is also added to the holograms. MATLAB is used to generate signals that synchronize the laser, SLMs, and camera. The lab setup, originally shown in Ref. (2), is displayed in Figure 4.1. The images generated by the free-space camera are 512x512 pixels and subsequently cropped to 256x256 pixels. They are resized to 128x128 for computational efficiency.

The free-space OAM images consist of 32 different patterns, created through multiplexing beams passed through combinations of five different phase plates with modes [-4, -1, 2, 5, 8].

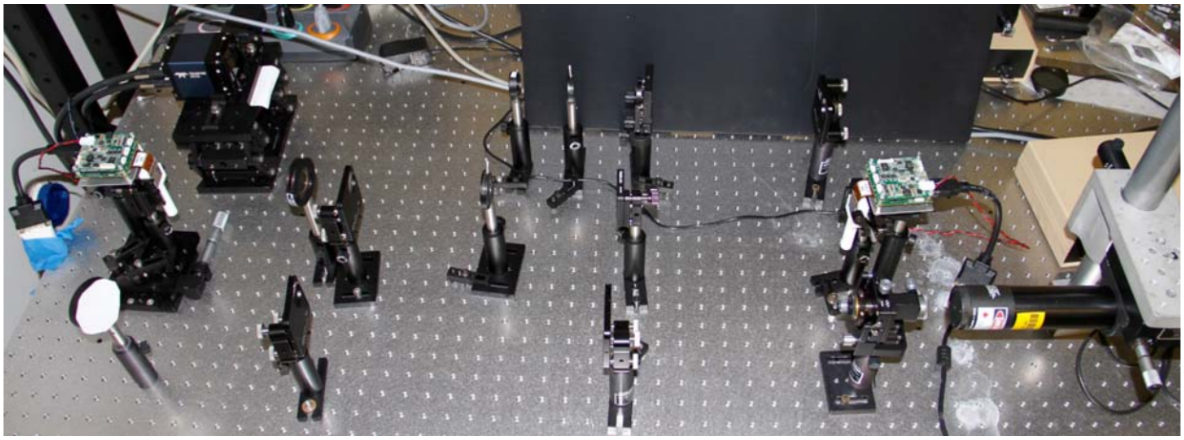


Fig. 4.1: Bench setup for free-space configuration.

OAM patterns from this set are shown in Figure 4.2.

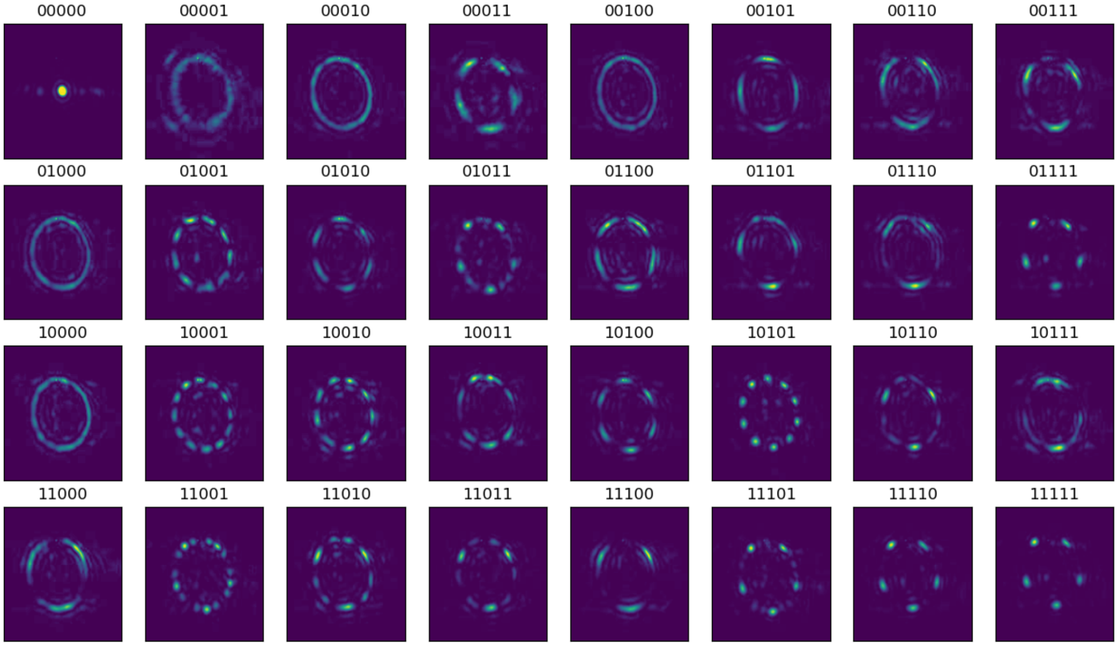


Fig. 4.2: Example of OAM patterns from the free-space data set.

The free-space set is composed of three distinct data collects, where each group is collected at a specific turbulence level. Turbulence is simulated and imparted through inserting phase screens in the OAM beam path. The turbulence levels are $D/r_0 = 5$, $D/r_0 = 10$, and $D/r_0 = 15$. Examples of OAM patterns at different turbulence levels are showing in Figure

4.3. The data collects will be referred to as TB5, TB10, and TB15, where each data collect contains images from each of the 32 patterns.

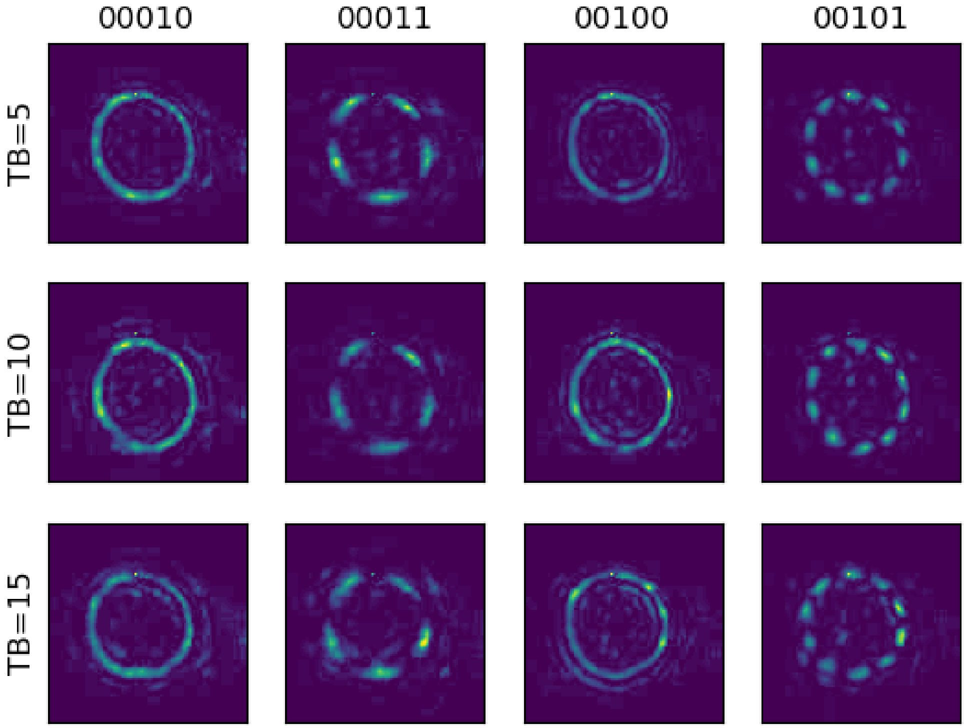


Fig. 4.3: Example of different turbulence levels from the free-space data set. Column header indicates the pattern number and the row label indicates the level of turbulence. Inspecting the different levels for the OAM modes shows pattern displacement or distortion due to the turbulence.

Underwater data is collected using the following hardware setup. The laser source used in these measurements was a diode pumped solid state laser that operates at 532-nm and produces 5 nS pulses with 250 uJ/pulse. The intensity patterns are captured by a high performance, fast-frame-rate camera (Photron FASTCAM SA-Z). The camera is synchronized with the laser pulses at a rate of 1kHz. As shown in Figure 4.4, the laser beam splits into four coherent beams and is expanded to pass through and fill vortex phase plates where an OAM phase is imparted to each beam. After leaving the phase plates, the beams are recombined using beamsplitters. The multiplexed OAM beam passes through a 1.2-meter water tank and is routed to the camera using mirrors. Polyamid beads are added to the water to introduce signal scattering while small pumps agitate the water to ensure that the particles remain in suspension and homogeneously distributed. Attenuation length is measured by running a 15 mW, 532 nm CW probe laser parallel to the beam that is incident on a sensor. The OAM modes used for this configuration are [1, 4, -6, -8]. Images created by the camera for this setup are 1024x1024 pixels. The images are cropped and then resized to 128x128 pixels.

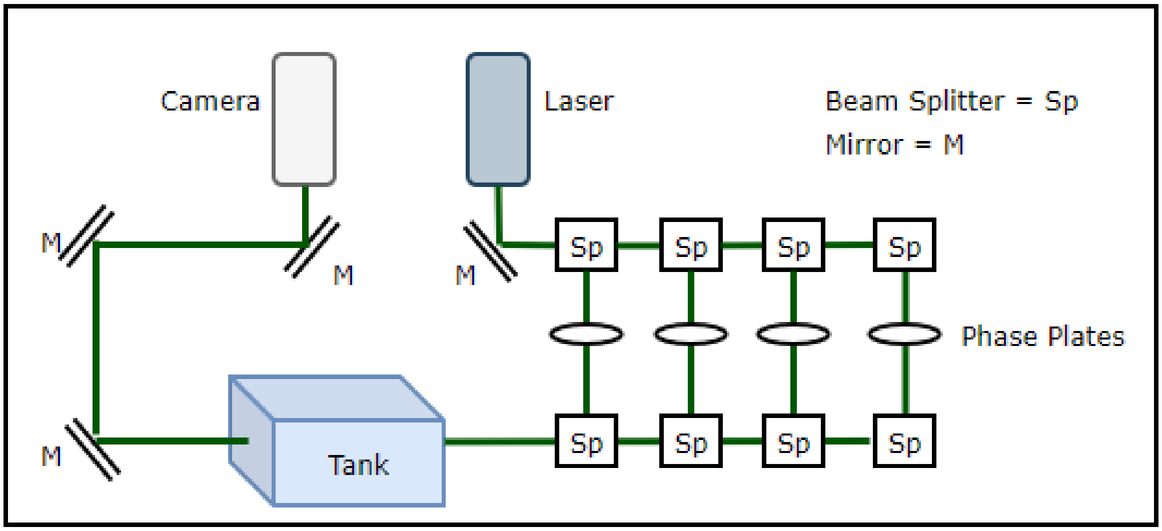


Fig. 4.4: Bench setup for underwater OAM communication configuration.

Underwater OAM images consist of 16 unique patterns derived from combinations of four different phase plates. Examples of the OAM patterns are shown in Figure 4.5.

The underwater set is composed of four different data collects, where each group is collected at a specific level of attenuation. Attenuation is created by adding polyamid beads designed for scattering light. Attenuation length sets are composed of levels 0, 4, 8, and 12. The sets will be referred to as AL0, AL4, AL8, and AL12. Each set contains images representing each of the 16 OAM patterns. Figure 4.6 shows attenuated patterns at the four levels of interest.

Each free-space and underwater data set is divided into training, validation, and test sets at a 70%/15%/15% respective split. The test sets are used only after a classifier is fully

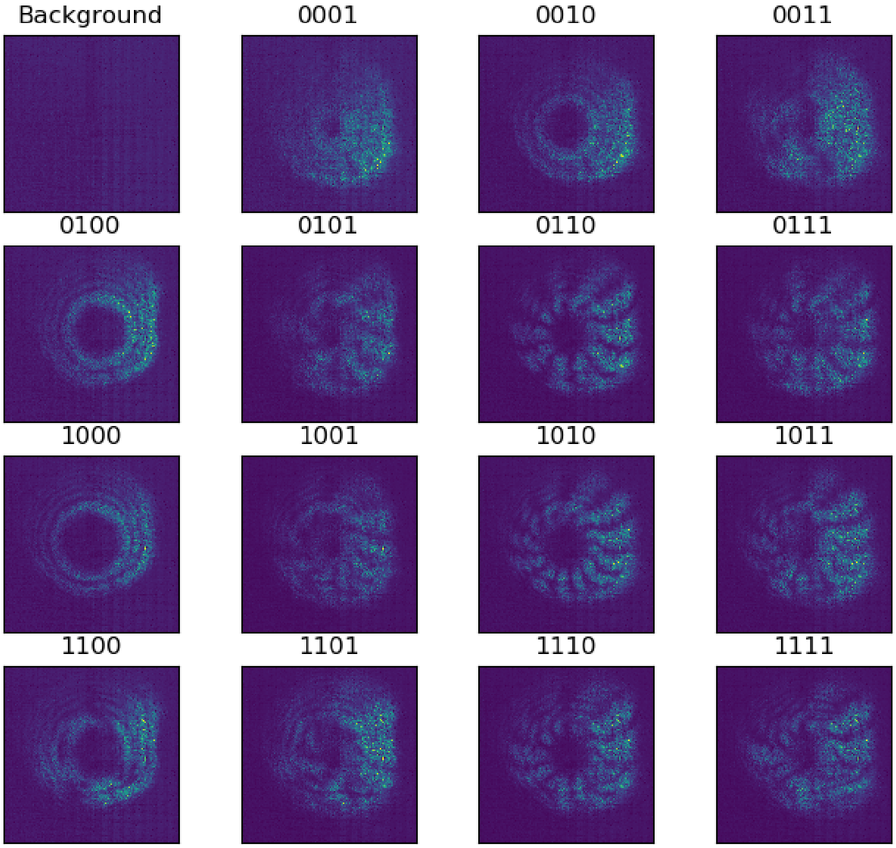


Fig. 4.5: Example of OAM patterns from the underwater data set.

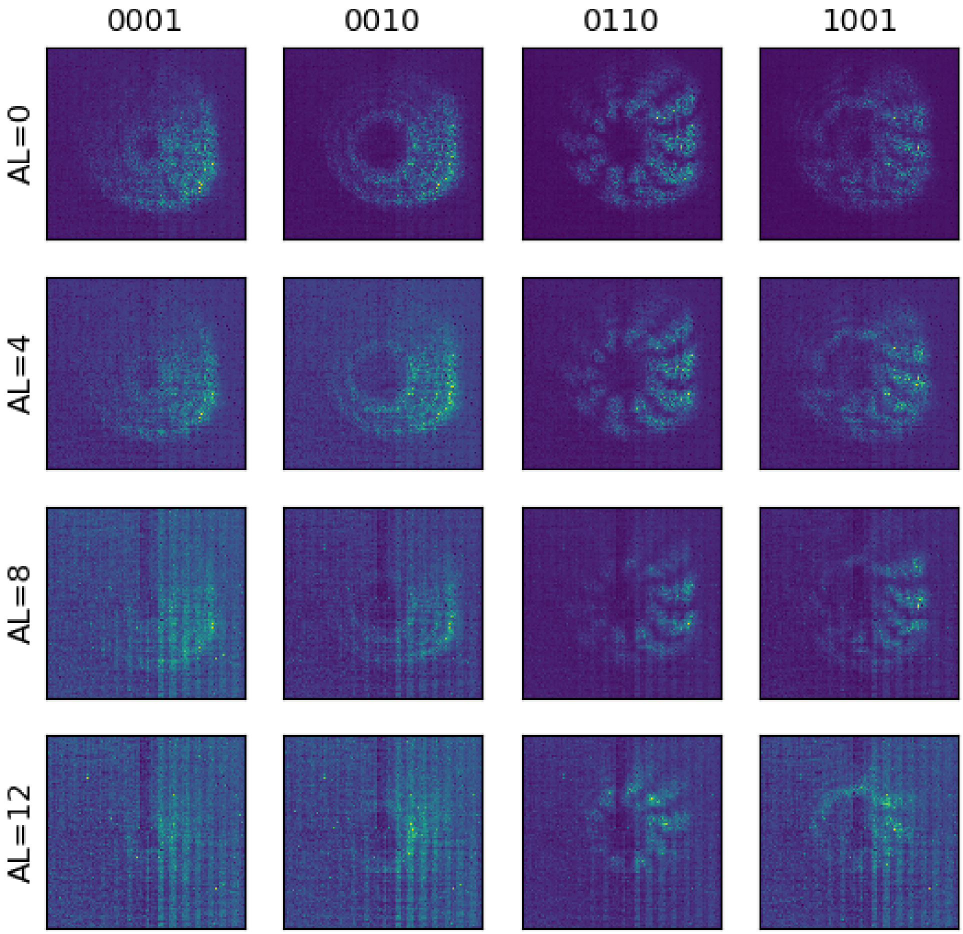


Fig. 4.6: Example of different attenuation levels from the underwater data set. Column header indicates the pattern number and the row label indicates the attenuation level.

trained and its final performance metrics are gathered. The test set is often referred to as a holdout set, as it is put aside until the very end.

Training took place on a computer with a NVIDIA RTX 2080 GPU with 8 GB of RAM. The computer also has 32 GB RAM and an Intel i9 processor with 16 cores. For the CNN training and evaluation in this paper, the code is developed in Python and the ML library used is PyTorch.

Table 4.1 provides the total number of trainable parameters for the architectures used in this study, as reported by PyTorch.

Table 4.1: **CNN Architecture Trainable Parameter Count**

CNN	Parameter Count
ShuffleNet	374592
SqueezeNet	751840
MobileNet	2264864
DenseNet	6982048
ResNeXt	23045472
ResNet	23573600
AlexNet	57134944
WideResNet	66899808

4.5 Results

In this section, results for each of the tests are presented. In Section 4.5.1, results are presented for the hyperparameter search in free-space and underwater environments for each of the CNN architectures. Section 4.5.2 presents results for architectures trained against each data set. Finally, Section 4.5.3 shows results when architectures are trained with one data set and tested against other, more distorted, data sets.

4.5.1 Hyperparameter Tuning

Architectures are initialized with pre-trained weights from ImageNet training, and are used as the starting point for training the OAM patterns. As the ImageNet competition has 1,000 classes and a fixed input size of 224x224, the CNN input and output layers were modified for 128x128 sized input images and output dimensions for classes of 16 (underwater) and 32 (free-space).

As hyperparameter searches can quickly become computationally expensive, a two-tier approach is taken to narrow the field. The first tier is to identify an optimizer that provides the best results. The second tier it to identify the best learning rate for each architecture using the selected optimizer. For the hyperparameter study, the ResNeXt 50 architecture is used, training is limited to 5 epochs, and the TB5 data set is used. Batch size is set to 32 for DenseNet because of its memory requirements during the training process. All other architectures use batch sizes of 128.

Optimizers are selected from the set of Adam, RMSProp, AdaMax, and Nadam. The ResNeXt 50 architecture is used to train data from the TB5 free-space data set. Learning rates are selected from a range of $1e-6$ to $1e-1$ for each optimizer. A quick random search was first performed to find a few good performing starting values for each optimizer. Those values are then used as best guesses to seed a Hyperopt search to support the optimizer analysis. The Hyperopt search is allowed to run 25 iterations to identify the best performing learning rate.

Figure 4.7 plots accuracies achieved using the four optimizers from learning rates selected by Hyperopt search. The x-axis shows the learning rates on a log scale, while the y-axis represents the accuracy achieved on the holdout set after 5 epochs of training. It is interesting to note that all of the optimizers achieve similar peak accuracies and the overall distribution of accuracies is very similar. The primary difference being the offset of the accuracy curves relative to the learning rate.

Figure 4.8 shows an accuracy curve for each optimizer over the course of 60 epochs. The TB5 data set is used for training the ReNeXt architecture in this figure. Learning rates for each optimizer are derived from the peaks from Figure 4.5. Figure 4.8 shows similar convergence rates for all of the optimizers.

As Figure 4.8 shows similar performance between the optimizers, a simple statistical analysis is employed to make the selection of which optimizer to use. Table 4.2 shows the average and standard deviations of accuracies for epochs 20-70 from Figure 4.8. Results in the

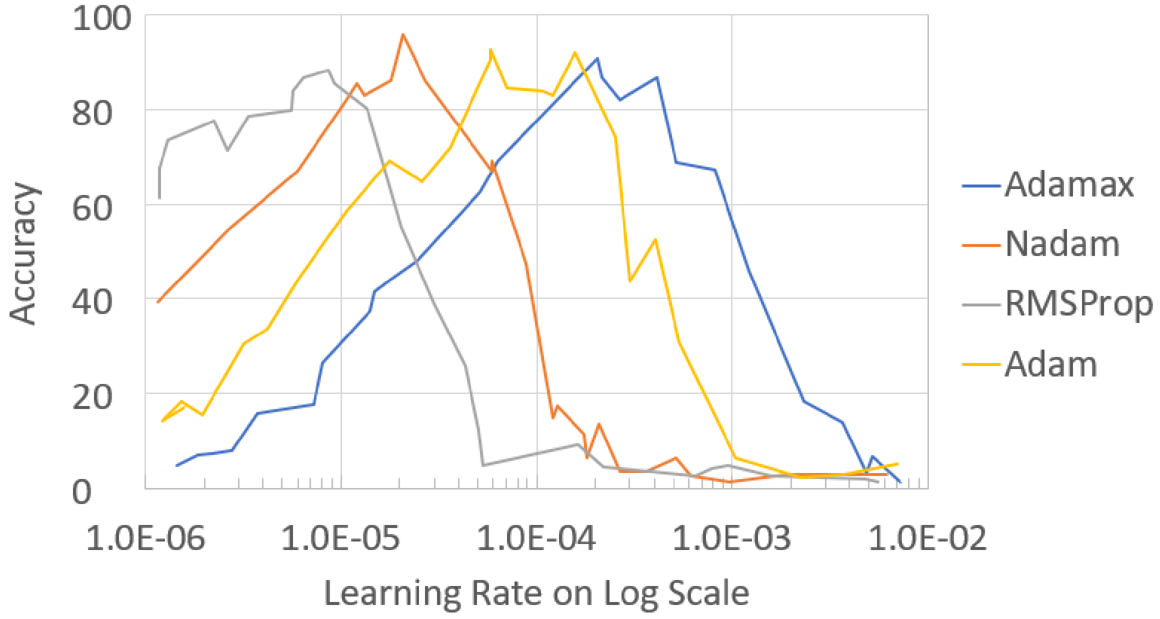


Fig. 4.7: Optimizer accuracy to learning rate comparison using TB5 free-space data set and ResNeXt 50 architecture. Accuracies are recorded after 5 epochs of training for learning rates selected by the Hyperopt algorithm.

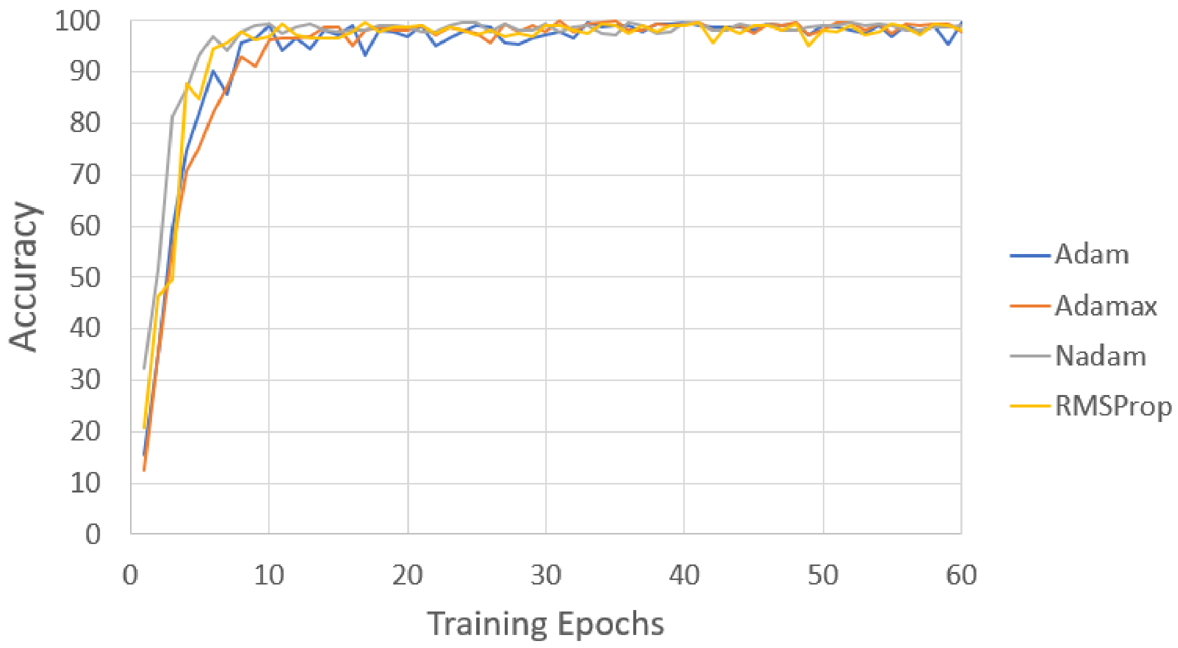


Fig. 4.8: Optimizer training curve comparison using TB5 free-space data set with ResNeXt 50 architecture. Accuracies are recorded after each training epoch for a total of 60 epochs.

table show that Nadam gets better average accuracy and lower standard deviation than the other optimizers. Consequently, Nadam is selected as the default optimizer for all subsequent training in this paper.

Table 4.2: **Optimizer Averages and Standard Deviations**

	Avg	SD
Adam	97.9	0.0140
Adamax	98.7	0.0085
Nadam	98.7	0.0064
RMSProp	98.3	0.0091

With the optimizer selected, the second tier of the hyperparameter search is to identify the best learning rates for each CNN architecture. Given that there are potential differences between the free-space and underwater data sets, this search is applied to each domain to see whether there are any significant differences in learning rate selection.

Figures 4.9 and 4.10 show Hyperopt results for accuracy vs. learning rate. The training is limited to 7 epochs, which is sufficient to generate curves showing relative training responses for different learning rates. The region of the figures of primary interest is the rising portion of the curve as these regions suggest the best place to draw learning rates from.

For the underwater set, Figure 4.9 shows very similar curves for the ResNet family of architectures. ShuffleNet shows the most difference as its learning rates are shifted to the right. Differences between the architectures are more pronounced in the free-space data as shown in Figure 4.10. Again, the ResNet family of architectures are similar at the same range of learning rates while ShuffleNet is also shifted far to the right in its learning rate curve. SqueezeNet appears to learn significantly slower than the other architectures. This graph turns out to be indicative of its overall performance later in the paper.

These curves provide an idea of what learning rate to use for training the architectures. Learning rates are selected moving from the left side of the curve (which begins at $1e-7$) and are selected at approximately 95% of the peak value. This allows selection of learning rates with good efficiency, but are not so high as to create convergence problems. This learning rate selection approach was established by Ref. (35).

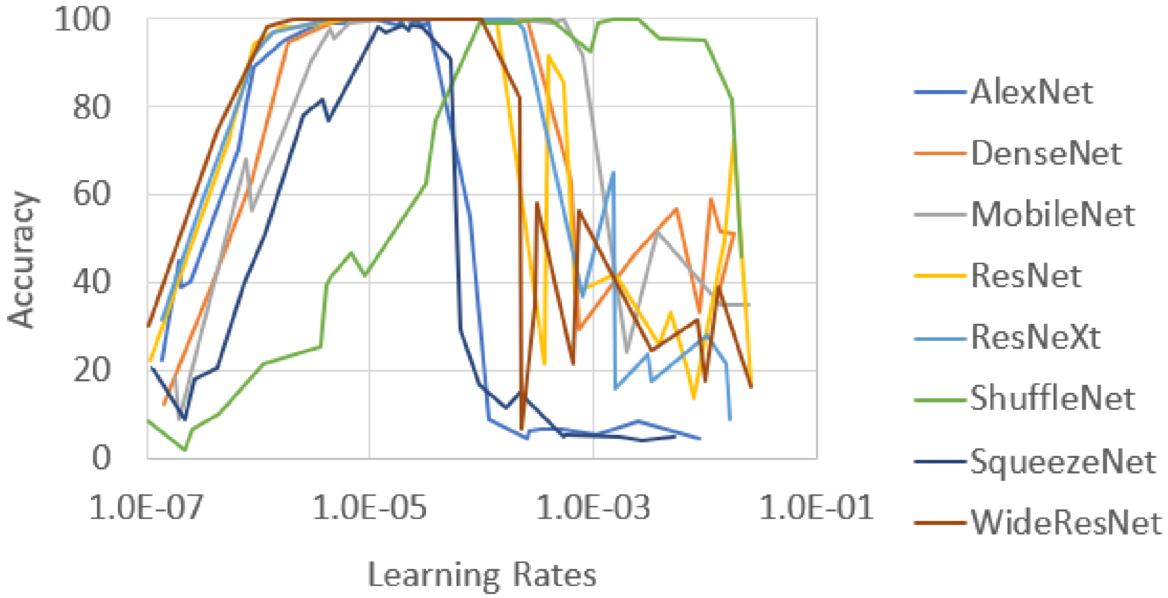


Fig. 4.9: Hyperopt learning rate search results for AL4 underwater image set.

Fig. 4.10: Hyperopt learning rate search using TB5 free-space image set.

Final learning rates used for each architecture, in underwater and free-space environments, are derived from these figures. Results are shown in Table 4.3. These are the learning rates used for the rest of the training in this paper. It is interesting to note that the learning rates between the two data sets are fairly similar to each other.

Table 4.3: **Final learning rates for architectures in underwater and free-space data sets.**

Architecture	Underwater	Free-space
AlexNet	1.2e-5	1.1e-5
DenseNet	4.2e-5	4.2e-5
MobileNet	1.2e-4	7.4e-4
ResNet	3.2e-5	3.2e-5
ResNeXt	2.4e-5	2.4e-5
ShuffleNet	4.2e-4	3.1e-4
SqueezeNet	1.5e-5	5.5e-5
WideResNet	1.4e-5	1.4e-5

Using the established learning rates, accuracy curves were generated for each architecture. This provides an initial comparison of how quickly the architectures learn and the levels that they converge to.

Figure 4.11 shows accuracy per training epoch curves for the underwater AL4 data set for each architecture (at the learning rates indicated in Table 4.3). It is apparent from the curves that, over time, all of the architectures achieve fairly similar accuracies.

Figure 4.12 shows accuracy per training epoch curves for the free-space TB5 data set for each architecture. Most of the architectures settle in at approximately the same end accuracy, the lone difference being SqueezeNet.

Aside from SqueezeNet, there does not appear, at this point, to be a great deal of difference from one architecture to another when using the AL4 (underwater) and TB5 (free-space) data sets. With hyperparameters selected, the architectures are ready for training against the data sets.

4.5.2 Baseline, Intra-set Tests

Baseline performance of the underwater data sets (AL0, AL4, AL8, AL12) and free-space

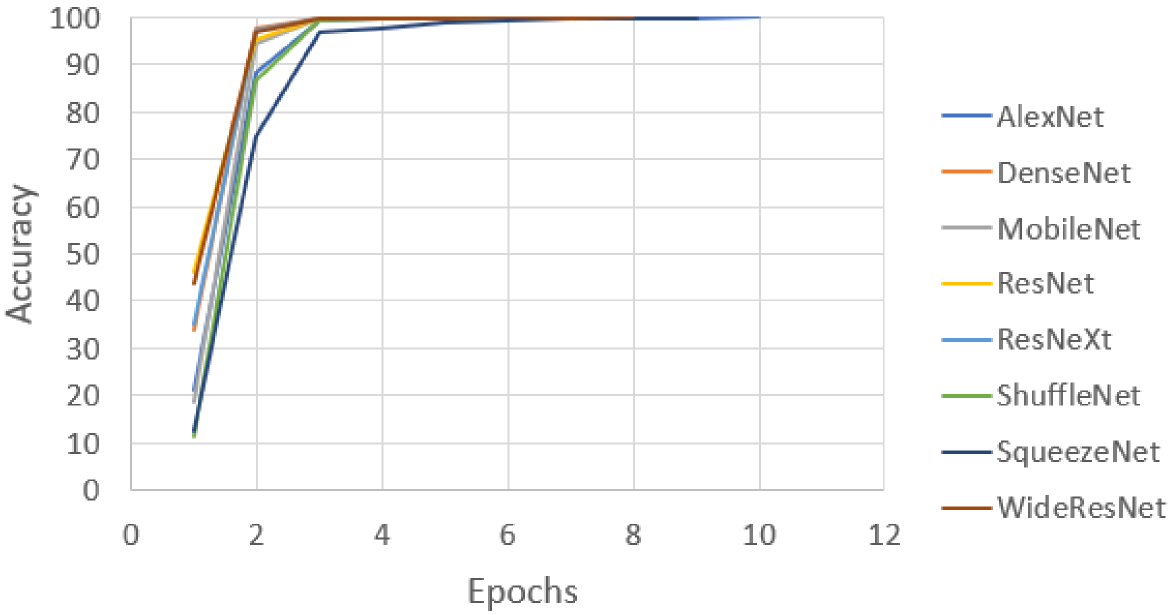


Fig. 4.11: Accuracy training curves for underwater data.

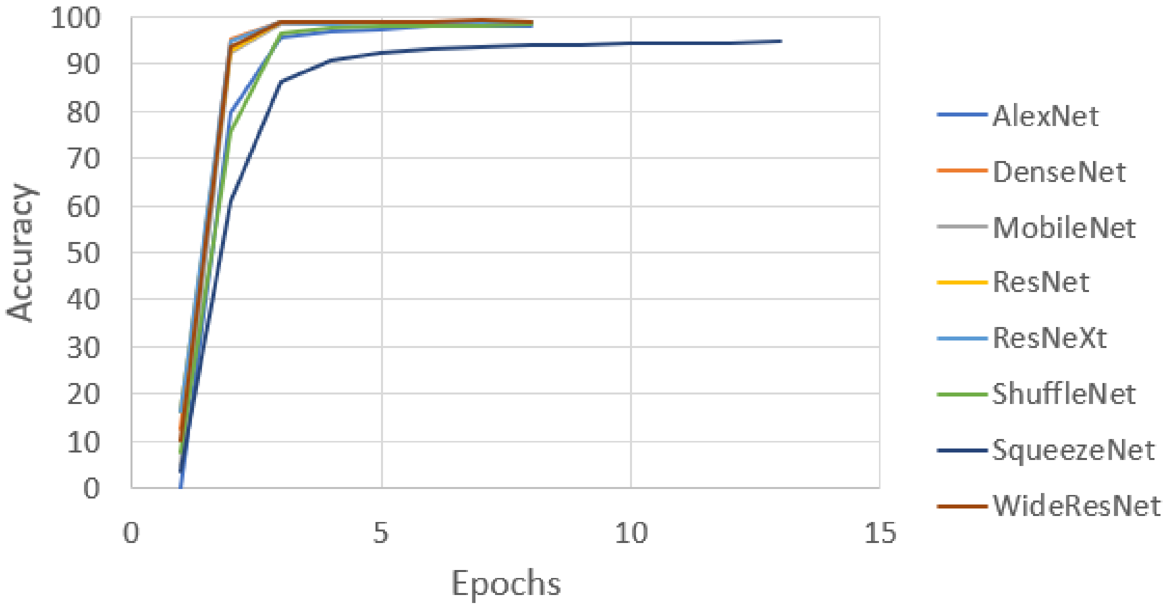


Fig. 4.12: Accuracy training curves for free-space data.

sets (TB5, TB10, TB15) is established in this section. In establishing baseline performance, the focus is placed on training an architecture with one data set and testing it the corresponding holdout set. Later, an architecture will be trained with one data set (TB5 for example) and then the holdout sets from TB10 and TB15 (inter-set testing) will be used to explore how well the architecture is able to generalize on data outside the training set. Percent accuracy is used at the metric for comparing relative performance of the different architectures.

Table 4.4 shows the baseline results for architectures trained with the underwater attenuated sets. As a whole, the architectures perform very well. Most accuracies achieved are 100%, or close to it. The only outlier is SqueezeNet on the AL12 set.

Table 4.4: **Architecture baseline performance with underwater sets.**

	AL0	AL4	AL8	AL12
SqueezeNet	99.5	99.4	99.4	98.8
AlexNet	99.4	100.0	99.4	99.4
ShuffleNet	100.0	99.4	100.0	100.0
DenseNet	100.0	100.0	100.0	99.4
ResNeXt	100.0	100.0	100.0	99.4
MobileNet	100.0	100.0	100.0	100.0
ResNet	100.0	100.0	100.0	100.0
Wide ResNet	100.0	100.0	100.0	100.0

Table 4.5 shows the baseline results for architectures trained with the free-space turbulent sets. This table includes a column that averages the results to help with sorting the architectures. The results in these tables are sorted by accuracy and show that the complexity imposed by turbulence, effects accuracy more than attenuation does. This table shows that AlexNet and SqueezeNet struggled the most. DenseNet appears to provide the best performance with the free-space data sets.

Table 4.6 shows an example of the amount of time it took to train each architecture. The table shows the number of training epochs as well as the amount of time it took to train for the specified number of epochs. The training loop had a maximum number of training epochs, but was allowed to terminate early when a specific level of accuracy (99.9%) had been achieved on the validation set. Most of the architectures trained quickly in terms of epochs and overall time. AlexNet and SqueezeNet took the longest amount of time to train while yielding the poorest results. DenseNet and the ResNet family required the fewest training epochs and took a comparable amount of time to train.

4.5.3 Inter-Set Performance Analysis

Section 4.5.2 shows that most of the architectures perform well when tested with the holdout test sets from the original data set. In real environments the trained classifiers are likely to be presented with images that have been distorted by larger turbulence and attenuation than what was present in the training set. The tests in this section explore the architectures and how well they perform when presented with images outside of their training set. For example, how well does an architecture trained with the AL0 data set classify attenuated images from the AL4, AL8, and AL12 holdout sets?

For this analysis, both underwater and free-space data sets are evaluated. For the underwater sets, the AL0 and the AL0-4 trained architectures are used. These architectures are evaluated against the AL0, AL4, AL8, and AL12 holdout sets. For the free-space data sets, the TB5 and TB5-10 trained architectures are used. These architectures are evaluated with the TB5, TB10, and TB15 holdout sets. In both cases, the results of interest are with data sets that fall outside the training sets. In the following tables, the results are ordered by ascending accuracies.

Table 4.7 contains results from architectures that have been trained with the AL0 data set. The four columns contain accuracies from AL0, AL4, AL8, and AL12 holdout sets. In looking at the performance of the AL4 test set, DenseNet and Wide ResNet give the best accuracies at 63.7% and 81.2% respectively.

In evaluating architectures trained with the combined AL0 and AL4 data sets, Table 4.8 shows results ordered according to ascending results for the AL8 data set. DenseNet and ShuffleNet take the lead spots with 80.0% and 84.4% respectively.

Tables 4.9 and 4.10 have similarly organized results for free-space data sets. ResNet and DenseNet (97.8% and 97.9%) have the best results for TB10 in Table 4.9, while DenseNet and ResNet (81.8% and 84.8%) take the lead spots in Table 4.10.

Table 4.5: **Architecture baseline performance with free-space sets.**

	TB5	TB10	TB15	Avg
SqueezeNet	96.9	90.4	88.3	91.9
AlexNet	99.7	99.6	83.0	94.1
ShuffleNet	99.9	99.9	100.0	99.9
MobileNet	99.8	100.0	100.0	99.9
Wide ResNet	99.8	100.0	100.0	99.9
ResNet	99.9	100.0	100.0	99.9
ResNeXt	99.9	100.0	100.0	99.9
DenseNet	100.0	100.0	100.0	100.0

Table 4.6: **Training epochs and time for the TB5 free-space set.**

Architecture	Epochs	Time (sec)
AlexNet	80	636
DenseNet	5	96
MobileNet	20	130
ResNet	6	110
ResNeXt	7	78
ShuffleNet	21	80
SqueezeNet	80	344
Wide ResNet	6	90

Table 4.7: **Underwater AL0 inter-set test. Architectures trained on ALO and tested against all AL data sets.**

	AL0	AL4	AL8	AL12
MobileNet	100.0	29.4	3.1	3.7
AlexNet	99.4	44.4	18.8	10.6
SqueezeNet	97.5	45.0	23.7	7.5
ShuffleNet	100.0	45.6	16.3	6.9
ResNet	100.0	46.9	13.8	15.6
ResNeXt	100.0	51.2	13.1	12.5
DenseNet	100.0	63.7	18.1	5.0
Wide ResNet	100.0	81.2	7.5	5.6

Table 4.8: **Underwater AL0-4 inter-set test. Architectures trained on ALO-4 and tested against all AL data sets.**

	AL0	AL4	AL8	AL12
ResNet	100.0	99.4	46.3	19.4
AlexNet	99.4	100.0	58.8	13.1
MobileNet	100.0	100.0	61.3	13.1
Wide ResNet	99.4	100.0	70.0	23.1
SqueezeNet	100.0	99.4	71.3	17.5
ResNeXt	100.0	99.4	79.4	36.9
DenseNet	100.0	100.0	80.0	40.0
ShuffleNet	100.0	100.0	84.4	45.0

Table 4.9: **Free-space TB5 inter-set test. Architectures trained on TB5 and tested against all TB data sets.**

	TB5	TB10	TB15
SqueezeNet	98.6	83.9	62.4
AlexNet	97.4	86.4	58.0
ShuffleNet	97.6	92.2	75.9
MobileNet	98.8	93.9	77.0
Wide ResNet	98.6	94.5	82.2
ResNeXt	98.9	95.4	81.4
ResNet	98.8	97.8	87.8
DenseNet	99.6	97.9	85.5

Table 4.10: **Free-space TB5-10 inter-set test. Architectures trained on TB5-10 and tested against all TB data sets.**

	TB5	TB10	TB15
AlexNet	97.4	82.0	59.8
SqueezeNet	98.6	83.2	61.1
MobileNet	97.6	91.3	74.0
Wide ResNet	98.6	95.1	75.2
ShuffleNet	99.0	93.5	78.5
ResNeXt	97.9	94.1	79.3
DenseNet	99.0	96.7	81.8
ResNet	98.9	96.9	84.8

4.6 Discussion

For this effort, three lightweight architectures (MobileNet, ShuffleNet, and SqueezeNet) and five heavyweight architectures (AlexNet, DenseNet, ResNet, ResNeXt, and Wide ResNet) are evaluated.

From Section 4.5.2, the free-space data set provides interesting insight in performance differences. Table 4.5 shows DenseNet performing the best, however the ResNet family of architectures follow very closely in achieved accuracies.

The main conclusion from the underwater results in Table 4.4 is to avoid the SqueezeNet architecture. All other architectures seem to give comparable performance.

Another observation in comparing Tables 4.4 and 4.5 is the difference in accuracies between the two. The underwater data sets have great performance, even with the high signal-to-noise ratio. Why do the CNNs perform so well with the attenuated versus the turbulent images? The answer is likely due to the fact that, the overall shape of the attenuated images remains constant, where turbulence causes displacement of the image patterns. The CNNs have to work harder to learn many potential patterns that belong to a specific OAM set. In addition, the turbulence patterns have twice the number of image patterns to learn.

Transitioning to out-of-training-set testing provides even further insight into architecture robustness. Tables 4.7, 4.8, 4.9, and 4.10 have one consistent top performer, DenseNet. DenseNet took about the same amount of time and number of epochs to train as the ResNet family. It was also smaller than the ResNet family of architectures by at least a factor of 3. So, it was a consistent top performer, while also having significantly fewer parameters than many of the close performing architectures.

For a resource constrained system that can not support the size of DenseNet or other heavy-weights, what is the next best performer? The options come down to ShuffleNet, SqueezeNet, and MobileNet. Interestingly, in reviewing tables from Sections 4.5.2 and 4.5.3, ShuffleNet consistently performed as well as or better than the other architectures most of the time. Additionally, it is considerably smaller than the other two architectures.

4.7 Summary and Conclusions

This paper set out to evaluate OAM transmitted turbulent, free-space and attenuated, underwater images. The specific purpose is to identify, which, if any, of the state-of-the-art CNN architectures performs best in the two environments.

Steps were taken, using a parameter search, to identify the best performing optimizer and near-ideal learning rates for training the architectures. Four underwater and three free-space data sets were then used to train each architecture.

With the trained architectures, test sets were presented from each data set to evaluate their classification accuracy. For the baseline testing (no out of set images), architectures with the underwater set appeared to perform at a level comparable to each other. The lone outlier was SqueezeNet. The turbulent free-space images, however, presented a challenge sufficient to start separating out varying levels of performance. DenseNet was the best performer from this set of tests.

Advancing to the inter-set testing, additional performance differences emerged. These tests took architectures trained with a data set, such as AL0, and then that trained architecture was presented holdout data from more attenuated data sets. This was repeated with the turbulent, free-space data as well.

In the end, it was found that DenseNet consistently performed the best, or as a close second, all the time. This finding is interesting considering its parameter count is at least a factor of three less than the other architectures that performed well. This result implies that DenseNet generalizes well during training.

For systems that are more resource constrained, three architectures with lower parameter count were tested: ShuffleNet, SqueezeNet, and MobileNet. In evaluating the results, ShuffleNet consistently performed better than or was close to the performance of the other two. It is also has considerably fewer parameters than the others. For a resource constrained system, ShuffleNet would be the clear choice due to its low parameter count and competitive accuracy.

For the turbulent, free-space data sets DenseNet was trained for 5 epochs with the learning rate set to $4.2e-5$, while ShuffleNet was trained for 21 epochs with the learning rate set to $3.1e-4$. Their respective batch sizes were 32 and 128. Nadam was used as the optimizer used in training all architectures.

For the attenuated underwater data sets, DenseNet was trained for 5 epochs with a learning rate of $4.2e-5$ and a batch size of 32. ShuffleNet was trained for 21 epochs with the learning rate set to $4.2e-4$ with batch size of 128.

This work provides an in-depth comparison of state-of-the-art, deep convolutional neural networks in turbulent free-space and attenuated underwater environments. It shows which architectures provide the most robust performance for degraded conditions outside of the training set. For low memory systems, ShuffleNet performed the best. For higher capacity systems, DenseNet consistently performed the best. Training parameters for both architectures are provided. Future work includes evaluating architectures in environments that include both attenuation and turbulence. Also, it may be that a better architecture, with fewer parameters still exists that is best suited for OAM images. A network architecture search could be used to identify a better architecture for classifying OAM images.

4.8 Acknowledgements

We would like to acknowledge support for this project from the US Naval Research Laboratory through a 6.1 base program for orbital angular momentum and a 6.2 base program for underwater communications. Thank you to the research group of Eric Johnson for fabrication of the vortex phase plates.

REFERENCES

[1] M. Krenn, R. Fickler, M. Fink, J. Handsteiner, M. Malik, T. Scheidl, R. Ursin, and A. Zeilinger, "Communication with spatially modulated light through turbulent air across vienna," *New Journal of Physics*, vol. 16, no. 11, p. 113028, nov 2014. [Online]. Available: https://doi.org/10.1088%2F1367-2630%2F16%2F11%2F113028

[2] T. Doster and A. T. Watnik, "Machine learning approach to OAM beam demultiplexing via convolutional neural networks," *Applied Optics*, vol. 56, no. 12, pp. 3386–3396, 2017.

[3] S. Lohani and R. T. Glasser, "Robust free space OAM communications with unsupervised machine learning," in *Frontiers in Optics + Laser Science APS/DLS*. OSA, 2019.

[4] S. Rostami, W. Saad, and C. S. Hong, "Deep learning with persistent homology for orbital angular momentum (OAM) decoding," *IEEE Communications Letters*, vol. 24, no. 1, pp. 117–121, 2020.

[5] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde, "Transport-based model for turbulence-corrupted imagery," *Applied Optics*, vol. 57, no. 16, pp. 4524–4536, 2018. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-57-16-4524

[6] B. Cochenour, K. Dunn, A. Laux, and L. Mullen, "Experimental measurements of the magnitude and phase response of high-frequency modulated light underwater," *Applied Optics*, vol. 56, no. 14, pp. 4019–4024, 2017. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-56-14-4019

[7] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, "Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes," *Physical Review A*, vol. 45, no. 11, pp. 8185–8189, 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.45.8185

[8] J. Wang, J.-Y. Yang, I. Fazal, N. Ahmed, Y. Yan, H. HUANG, Y. Ren, Y. Yue, S. Dolinar, M. Tur, and A. Willner, "Terabit free-space data transmission employing orbital angular momentum multiplexing," *Nature Photonics*, vol. 6, pp. 488–496, 07 2012.

[9] G. Gibson, J. Courtial, M. J. Padgett, M. Vasnetsov, V. Pas'ko, S. M. Barnett, and S. Franke-Arnold, "Free-space information transfer using light beams carrying orbital angular momentum," *Optics Express*, vol. 12, no. 22, pp. 5448–5456, Nov 2004. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-12-22-5448

[10] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger, "Entanglement of the orbital angular momentum states of photons," *Nature*, vol. 412, no. 6844, pp. 313–316, 2001. [Online]. Available: http://www.nature.com/articles/35085529

[11] M. P. J. Lavery, F. C. Speirits, S. M. Barnett, and M. J. Padgett, "Detection of a spinning object using light's orbital angular momentum," *Science*, vol. 341, no. 6145, pp. 537–540, 2013. [Online]. Available: https://science.sciencemag.org/content/341/6145/537

[12] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial, "Measuring the orbital angular momentum of a single photon," *Physical Review Letters*, vol. 88, no. 25, p. 257901, 2002.

[13] M. P. J. Lavery, G. C. G. Berkhout, J. Courtial, and M. J. Padgett, "Measurement of the light orbital angular momentum spectrum using an optical geometric transformation," *Journal of Optics*, vol. 13, no. 6, p. 064006, apr 2011. [Online]. Available: https://doi.org/10.1088%2F2040-8978%2F13%2F6%2F064006

[14] M. S. Soskin, V. N. Gorshkov, M. V. Vasnetsov, J. T. Malos, and N. R. Heckenberg, "Topological charge and angular momentum of light beams carrying optical vortices," *Physics Review A*, vol. 56, pp. 4064–4075, Nov 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.56.4064

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386

[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.

[18] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[20] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556 [cs]*, Sep. 2014, arXiv: 1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556

[21] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987–5995.

[22] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[23] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size,” *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks

for mobile vision applications," 2017, cite arxiv:1704.04861. [Online]. Available: http://arxiv.org/abs/1704.04861

[25] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6848–6856.

[26] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?" *CoRR*, vol. abs/1805.12177, 2018. [Online]. Available: http://arxiv.org/abs/1805.12177

[27] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The marginal value of adaptive gradient methods in machine learning," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4151–4161.

[28] S. Ruder, "An overview of gradient descent optimization algorithms," *CoRR*, vol. abs/1609.04747, 2016. [Online]. Available: http://arxiv.org/abs/1609.04747

[29] T. Dozat, "Incorporating nesterov momentum into adam," 2016. [Online]. Available: https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ

[30] T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude," COURSERA: Neural Networks for Machine Learning, 2012.

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2017.

[32] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal Machine Learning Research*, vol. 13, no. null, p. 281–305, Feb. 2012.

[33] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13. JMLR.org, 2013, p. I–115–I–123.

[34] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, “Tune: A research platform for distributed model selection and training,” *arXiv preprint arXiv:1807.05118*, 2018.

[35] L. N. Smith, “Cyclical learning rates for training neural networks,” *arXiv:1506.01186 [cs]*, 2015. [Online]. Available: http://arxiv.org/abs/1506.01186

CHAPTER 5

Conclusions

OAM optical communications provide the potential to alleviating bandwidth and communication infrastructure issues. While there is appreciable potential with this technology, there are serious hurdles to overcome before its adoption can be realized. Two fundamental issues arise through signal degradation from the environment, namely turbulence and attenuation.

While turbulence in the free-space domain is being actively investigated, little has been explored in underwater environments. The research in this dissertation embarks on that path and looks at how attenuation impacts communication performance.

The research started in Chapter 2, by providing a physics-based justification for extending optimal transport from free-space turbulence to an underwater attenuated environment. Previous research established the use of machine learning and optimal transport in the free-space domain. This paper justified the use of the optimal transport in the underwater domain. It then extended machine learning approaches from the free-space domain to underwater and established a baseline of performance.

Chapter 3 explored the use of SMART Models. A SMART model is a model that captures the underlying physics behind attenuation. SMART Models were trained to create a mapping of images, or intensity patterns, from a high SNR environment to a low SNR environment. The SMART Model was then inserted in the training pipeline for a CNN-based classifier. The SMART Model randomly selects images and applies an attenuation based transformation to the image. This allows retraining of the architecture using only one data set and simply updating the SMART Model to reflect the problem physics. This approach provides a powerful way to update models when deployed systems are exposed to environments that degrade their performance. Using the SMART Model allows capture of the new conditions, update to the SMART Model, and then retraining of the classifier with the updated model in the pipeline. Two types of models were explored.

A baseline of performance was established by training an architecture without the SMART Model in the training loop. That architecture was then tested against attenuated data that wasn't part of its training set. With that baseline established, two SMART Models were created and used while re-training architectures. The CNN-based SMART Model showed great improvement over the baseline. However, its performance was superseded by the R-CDT-based SMART Model. Interestingly the R-CDT-based SMART Model required only one example from each class. The R-CDT-based SMART Model provides a viable way to retrain architectures as the number of samples from the attenuated environment is negligible compared to standard CNN training set sizes.

Chapter 4 provided and in-depth exploration of the state of the art CNNs in both free-space and underwater OAM communications. Because grabbing the latest and greatest CNN doesn't guarantee the best performance, an evaluation of some of the top performing architectures was done, and then compared against the baseline architectures from previous works. This study concluded that the best architecture for both environments with their associated environmental problems was DenseNet. The study also looked at architectures with lighter memory footprints for robotic or embedded scenarios. It was shown that ShuffleNet provided the best performance from this class of architectures.

Moving forward, there are many questions that need to be addressed. One is exploration of the impact of changing distances between source and receivers. As images spread out over distance, what is the impact of patterns shifting in size? There are hardware considerations in keeping sources and receivers aligned for platforms that move in between communications. Studies of the best OAM modes to use, multiplexed modes that are most robust to turbulence and attenuation. Ideal information encoding schemes. These are just a few of the areas that still need to be addressed for OAM to become a viable means of communication.

APPENDICES

APPENDIX A

OAM Hardware and Data Set Details

A.1 OAM Overview

When light is generated by a laser, the light leaving the laser is collimated. In three dimensional space, these synchronized waves appear as plane waves. In addition to forward momentum light can also have angular momentum. In the 1930's Beth showed mechanical torque could be induced by light (1). In the 1990s, Allen et al. combined the idea of an optical vortex with OAM (2) and brought the idea back into the spotlight. If pushing on a door is analogous to the standard forward momentum, or radiation pressure, then orbital angular momentum could be thought of as twisting the door knob. This rotational force of the electric and magnetic fields is thought of propagating along the z-axis of a three dimensional coordinate system. The plane perpendicular to the z-axis is referred to as the transverse plane.

Orbital angular momentum can be represented in three dimensions by a helical shape rotating about the z-axis. A topical charge or mode number is defined for the wavefront that describes the number of rotations undergone in one wavelength. In the literature the topical charge is interchangeably referenced as l or m. The higher the topical charge, the faster the rotation around the z-axis.

When projected onto a surface, the wavefronts form rings whose diameter is related to the charge number m. The charge number can take on both positive and negative values and indicates the direction of rotation about the z-axis. Due to the twisting, the light waves at the center can cancel each other out, resulting in a dark spot, and is called an optical vortex or an optical void.

Another characteristic of OAM beams of different mode numbers is that they're completely orthogonal (2). Consequently, OAM beams can be multiplexed together, as shown in Figure A.1. Here the laser beam is generated and then split into four different beams. Each beam is incident on a phase plate where an OAM mode is imparted. The resulting OAM beams are multiplexed back together using the beam splitters and passed through a tank. The resulting beam is then bounced off of a few mirrors and incident on the camera.

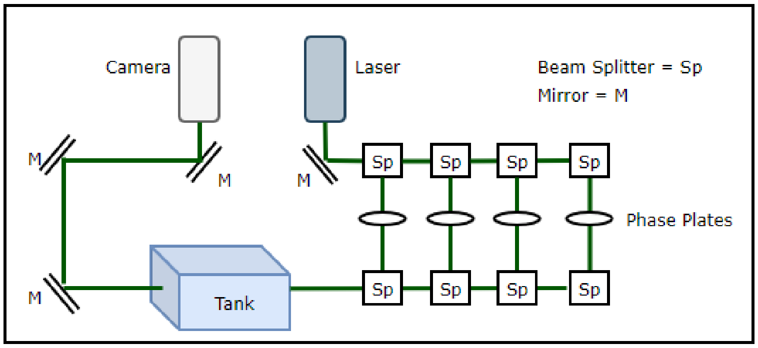


Fig. A.1: Examples of an underwater OAM beam bench top setup.

The multiplexed beams can take on complex patterns. An example of what the patterns may look like is shown in Figure A.2. These images were generated in the lab at the Naval Research Laboratory and were used in this research.

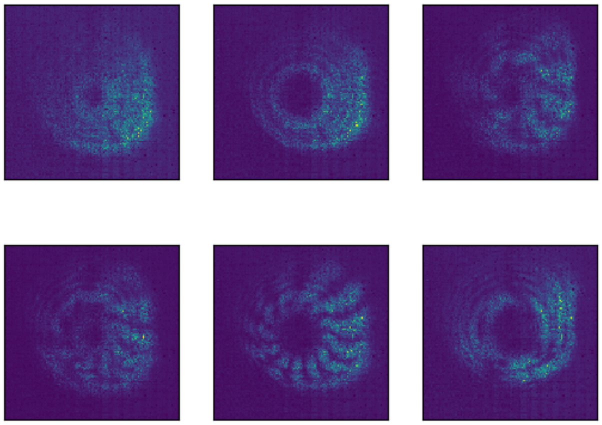


Fig. A.2: Examples of OAM patterns from an underwater OAM set. OAM images sampled from set produced from mode numbers [1, 4, -6, -8].

One characteristic of OAM patterns that may be observed, is how the patterns themselves may rotate in the transverse plane. If minor vibrations in the z-direction are present when recording light incident on the camera focal plane, then the pattern may appear to rotate. This is due to the helical path traveled by the electric field. If a transverse plane is moved back and forth along the z-axis and we're effectively taking a cross-sectional slice of that field, then one can conceptually envision why the image will rotate. Figure A.3 shows an example of several OAM patterns at different points in time and variation in their intensity patterns.

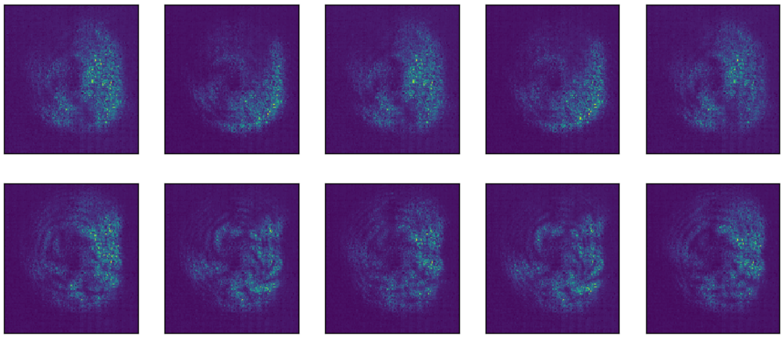


Fig. A.3: Examples of variations in OAM intensity patterns. Each row contains a single OAM mode combination. Each column is a different sample of the OAM mode in time.

OAM Mode Number Origin

A common reference in OAM literature is the ubiquitous reference to a phase number expressed as $exp(il\phi)$, where l referred to as a mode number. It turns out that not all lasers exhibit OAM properties. The beams that do, include Bessel (3), Bessel-Gauss (4), Laguerre-Gauss (2), Hermite-Gauss (5), Ince-Gauss (6), and Mathieu-Gauss (7).

A Gaussian beam is a special electromagnetic field whose EM field amplitudes are given by a Gaussian function. For most lasers, this means that the intended output displayed on a surface is a high intensity region in the middle with the intensity falling off as a Gaussian curve at the radius.

The Laguerre-Gaussian beam that sparked OAM interest, is a complete set of solutions to the paraxial Helmholtz equation. It is derived by writing the paraxial Helmholtz equation in cylindrical coordinates (ρ, ϕ, z), where ρ is the radius, ϕ is the angle in the plane perpendicular to the z axis, and z is the axis of propagation. The Laguerre-Gauss beam is described by Equation A.1

$$\begin{aligned} U_{l,m}(\rho,\phi,z) = A_{l,m}[\frac{W_0}{W(z)}](\frac{\rho}{W(z)})^l L_m^l(\frac{2\rho^2}{W^2(z)})exp(-\frac{\rho^2}{W^2(z)}) \\ \times exp[-jkz - jk\frac{\rho^2}{2R(z)} \mp jl\phi + j(l+2m+1)\zeta(z)], \end{aligned} \tag{A.1}$$

where $A_{l,m}$ is a constant, W_0 is the waist radius, $W(z)$ is the beam width, L_m^l represents generalized Laguerre polynomials, $R(z)$ is the radius of curvature for the wavefront at position z, and ζ is phase delay of the wavefront.

The OAM dependency is given by $exp(\mp j\ell\phi)$, which controls whether the EM wave experiences a helical wavefront ($|\ell| > 0$) or a plane wave ($\ell = 0$). The sign determines whether the twist is right or left.

The OAM dependency just described is a term that is frequently cited in papers. Background on where the value comes from is often assumed prior knowledge.

A.1 Orbital Angular Momentum - Imparting Methods and Characteristics

Imparting OAM modes to beams can currently be accomplished in a variety of ways. Generally a beam is incident on a surface with special characteristics. Surfaces that can impart OAM include computer-generated holograms (8), spiral phase plates (9), or cylindrical lenses (10). These devices effectively receive a plane wave at one surface and induce a twist such that an OAM wave comes out. Each OAM device is specially designed to impart a specific mode number and a specific direction to the incident beam. Practically, the number of modes is limited by physical constraints such as aperture size. The OAM beams for this research were generated using spiral phase plates to impart OAM modes.

As a side note, OAM isn't unique to only light waves. Since the original paper, these principles have been applied to sound waves (11), Doppler shifts (12), and radio waves (13). A nice summary of current work is here (14).

A.2 OAM Hardware Setup

All of the OAM data used for research in this dissertation was generated at the Naval Research Laboratory. The following information is available for anyone interested in creating a hardware setup like theirs for generating similar images. All of the data sets were created following a similar format and is detailed in Section A.3.

The laser source used in these measurements was a diode pumped solid state laser (Bright Solutions ONDA 532) that operates at 532 nm and produces 5 nS pulses with 300 uJ/pulse. The laser is externally triggered at 1 kHz using a waveform generator (Agilent 33600A) and the output beam was expanded to fill the aperture of the static vortex phase plates. The output beam was then split into four equal intensity beams and each beam was normally incident on one of the four phase plates of different charge [1, 4, -6, -8], imprinting a different OAM phase on each beam.

The phase plates were fabricated at Clemson University. The beams were then coherently recombined using beam splitters and steered through a 1.2 m water tank. The transmitted intensity patterns were captured by a high performance, fast-frame-rate camera (Photron FASTCAM SA-Z) that was externally triggered at the rate of 1000 frames per second, synchronized with the laser pulses.

One thousand images from each of the 16 permutations of the OAM beam set were captured for the analysis. The 16 OAM mode patterns were manually selected by blocking one or more of the beams. The camera spatial resolution is 1024x1024 pixels with a 20 m pitch and a bit depth of 12 bits.

Five um polyamid seeding particles (Dantec Dynamics) are added to the water in the tank to introduce attenuation into the beam path. The beam attenuation is continuously monitored by measuring the transmission of a 15 mW, 532 nm CW laser beam propagating through the tank parallel to the OAM beam. Small pumps agitate the water and insure that the particles remain in suspension and homogeneously distributed.

In previous work (15) ferroelectric spatial light modulators (SLMs) were used for generating different phase patterns. Since SLMs are relatively slow, specially fabricated phase plates were created to improve throughput rates.

A.3 OAM Data Set Details

Plastic beads with, five micron diameter, were added to water to attenuate signal intensity. The beads are designed scatter light and the amount of scattering is proportional to the number of beads present. A total of four data sets were made, one each for attenuation lengths 0, 4, 8, and 12. These data sets are referred to as AL0, AL4, AL8, AL12 in the research chapters. Signal attenuation lengths are described by Equation A.2, where i is the attenuated signal, i_0 is the original signal strength, and AL is the level of attenuation.

$$i = i_0 e^{-AL} \tag{A.2}$$

Figure A.4 shows an example of six classes at each of the attenuation lengths. In OAM pattern 0001, the AL0 image shows a distinct pattern with a high SNR. As the attenuation length increases, the noise floor of the camera moves up and the SNR drops. At AL12 we see the noise overcoming the lower intensity portions of the OAM pattern.

For training and testing, the AL sets were split into 70%/15%/15% for training, testing, and validation, respectively. Test sets were used only once, at the end to produce final results. Images from the camera were originally 1024x1024 grayscale images. The borders were cropped to remove pixels without intensity information and were downsampled to a final dimension of 128x128.

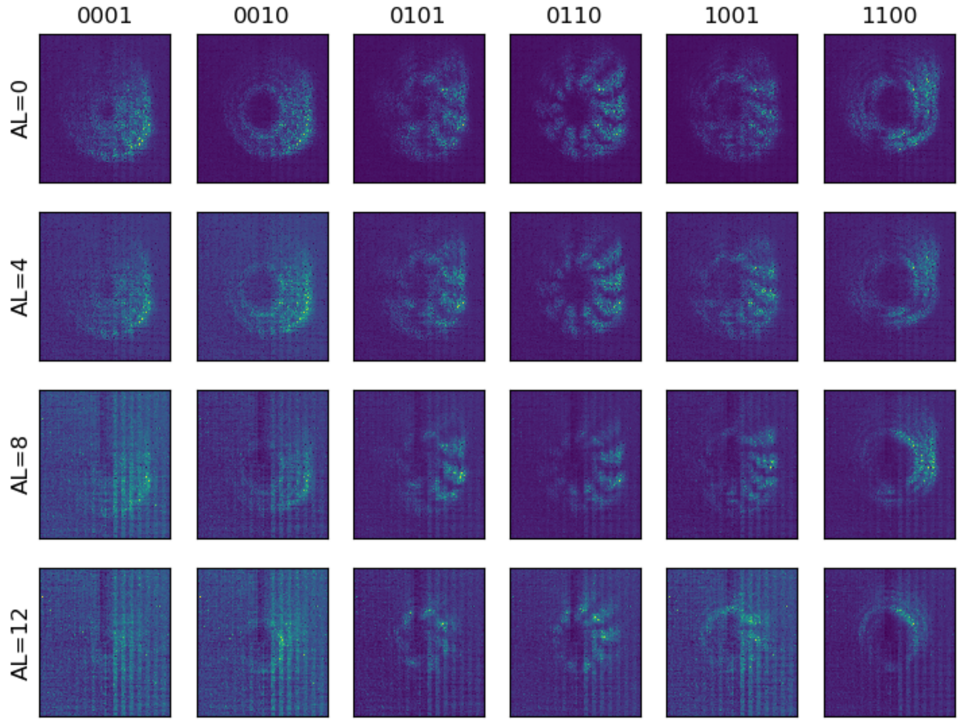


Fig. A.4: Examples of attenuation in underwater environment. Each column contains patterns for one OAM mode and each row represents a specific level of attenuation. The first row contains attenuation level 0. The subsequent rows show progressively lower SNR levels.

APPENDIX B

Optics Background

The motivation for the following derivations, is to follow the path in understanding how light (electromagnetic waves) are able to go from a plane wave to following a helical path. Also, much of the current literature on orbital angular momentum cite the expression '$\exp(i\ell\theta)$'. The path to this important expression is also illuminated.

In the following sections Stokes' Theorem will be reviewed and its impact on Maxwell's equations relating to electrical and magnetic fields will also be highlighted. With relationships from Maxwell's equations established, we'll explore Laguerre-Gaussian (LG) beams and the special properties that they possess. Finally, the relationship of LG beams and orbital angular momentum will be shown.

B.1 Stokes' Theorem

In vector calculus, one might be interested in measuring the the amount of a substance that flows through a cross-sectional region of interest. The quantity of substance (the measured quantity being defined as flux) might be a fluid such as water through a pipe or the amount of electricity flowing through a wire.

Across some surface '**S**' there may be a closed path '**C**' through which we are interested in evaluating the flux of some vector field '**F**'. The integral describing this measurement is shown in Equation B.1.

$$\int_C F ds \tag{B.1}$$

Stoke's Theorom states that we can divide the enclosed path into two paths, take their integrals, add them together, and have the same total flux as if we'd just used the single integral. This is represented by Equation B.2.

$$\int_C Fds = \sum_i \int_{C_i} Fds \tag{B.2}$$

If we continue taking smaller and smaller areas, we eventually wind up with a point in the area with a vector coming out of it. When the area has been reduced to a point, the local value is referred to as the *curl* and is denoted as $\bigtriangledown \times F$. As we're primarily concerned with the normal component of the vector at the specified point, the equation at that location is shown in Equation B.3.

$$(\bigtriangledown \times F) \cdot n = \lim_{a_i \ to\infty} (\frac{1}{a_i} \int_{C_i} Fds) \tag{B.3}$$

To get the integral of all curls over the surface of interest we get Equation B.4.

$$\int_S \bigtriangledown \times Fda \tag{B.4}$$

But, from Equation B.1 we have that the sums of all curls is nothing more than:

$$\int_C Fds \tag{B.5}$$

Stokes' Theorem says that Equations B.4 and B.5 are equal to each other. Hence we arrive a the Equation B.6:

$$\int_C Fds = \int_S \bigtriangledown \times Fda \tag{B.6}$$

This equation allows us to either compute a line integral over a surface boundary or the curl of a vector field. The two calculations are equivalent to each other and whichever is most convenient may be used.

Stoke's Theorem and Maxwell's Equations

Maxwell's equations play an important role in understanding electromagnetic propagation. Stoke's theorem helps to make some important generalizations in working with these equations.

From Ampere's Law we have that

$$\int_{loop} Bds = \mu_0 I_e ncl. \tag{B.7}$$

This equation tells us that the magnetic flux over a defined loop is proportional to the current through a similarly closed region (like a wire) times the vacuum permeability. Permeability expresses how easy it is for the magnetic flux to move through a medium.

Stokes' Theorem can be applied to this equation to relate the magnetic flux over the loop to the local curl such that:

$$\int_{loop} Bds = \int_{surface} \bigtriangledown \times Bda. \tag{B.8}$$

Ampere's law also states that:

$$\int_{loop} Bds = \mu_0 I_e ncl = \mu_0 \int_{surface} Jda. \tag{B.9}$$

Bringing the μ_0 inside the integral and setting equations equal to each other we get:

$$\int_{surface} \bigtriangledown \times Bda = \int_{surface} \mu_0 Jda. \tag{B.10}$$

From these integrals we can deduce that the integral contents can be set equal to each other:

$$\bigtriangledown \times B = \mu_0 J. \tag{B.11}$$

Another way this expression can be presented is:

$$\bigtriangledown \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{d}{dt} E. \tag{B.12}$$

A similar set of steps can be taken to deduce Faraday's law from:

$$\int_{loop} Eds = -\frac{d}{dt}\int_{S} Bda. \tag{B.13}$$

to:

$$\bigtriangledown \times E = -\frac{d}{dt}B. \tag{B.14}$$

This shows the relationship between changing magnetic flux and electric fields.

B.2 Deriving the wave equations from Maxwell's equations

Working with Maxwell's equations a little more we make the simplification of an electromagnetic wave propagating through a vacuum and charge free space. With these settings we have:

$$\bigtriangledown \cdot E = 0 \tag{B.15}$$

$$\bigtriangledown \times E = -\frac{d}{dt}B. \tag{B.16}$$

$$\bigtriangledown \cdot B = 0 \tag{B.17}$$

$$\bigtriangledown \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{d}{dt}E. \tag{B.18}$$

If we take the curl of Equations B.16 and B.18 we deduce:

$$\bigtriangledown \times (\bigtriangledown \times E) = \bigtriangledown \times (-\frac{dB}{dt}) = -\frac{d}{dt}(\bigtriangledown \times B) = -\mu_0 \epsilon_0 \frac{d^2 E}{dt^2}. \tag{B.19}$$

$$\bigtriangledown \times (\bigtriangledown \times B) = \bigtriangledown \times (\mu_0 \epsilon_0 \frac{dE}{dt}) = \mu_0 \epsilon_0 \frac{d}{dt}(\bigtriangledown \times E) = -\mu_0 \epsilon_0 \frac{d^2 B}{dt^2}. \tag{B.20}$$

Given the vector identities:

$$\bigtriangledown \times (\bigtriangledown \times \boldsymbol{V}) = \bigtriangledown(\bigtriangledown \cdot \boldsymbol{V}) - \bigtriangledown^2 \boldsymbol{V} \tag{B.21}$$

$$\bigtriangledown^2 \boldsymbol{V} = \bigtriangledown \cdot (\bigtriangledown \boldsymbol{V}) \tag{B.22}$$

The electrical field, **E**, and the magnetic field, **B**, can be substituted in for the vector fields, **V** of Equation B.21. Given the relationship between Equations B.19 and B.21 we can substitute the vector and values of B.19 into B.21 to get:

$$\bigtriangledown \times (\bigtriangledown \times \boldsymbol{V}) = \bigtriangledown(\bigtriangledown \cdot \boldsymbol{V}) - \bigtriangledown^2 \boldsymbol{V} \tag{B.23}$$

$$\bigtriangledown \times (\bigtriangledown \times \boldsymbol{E}) = \bigtriangledown(\bigtriangledown \cdot \boldsymbol{E}) - \bigtriangledown^2 \boldsymbol{E} \tag{B.24}$$

We know that $\bigtriangledown \cdot E$ is 0 by Equation B.15. Substituting into B.24 we get:

$$\bigtriangledown \times (\bigtriangledown \times \boldsymbol{E}) = - \bigtriangledown^2 \boldsymbol{E} \tag{B.25}$$

Substituting Equation B.19 into B.25 we get:

$$-\mu_0 \epsilon_0 \frac{d^2 E}{dt^2} = - \bigtriangledown^2 \boldsymbol{E} \tag{B.26}$$

Given that the speed of light is $c_0 = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$ and rearranging the sides in Equation B.26 we arrive at:

$$\frac{1}{{c_0}^2} \frac{d^2 \boldsymbol{E}}{dt^2} - \bigtriangledown^2 \boldsymbol{E} = 0 \tag{B.27}$$

Following a similar process of substitutions we can arrive at a similar wave equation for magnetic fields:

$$\frac{1}{{c_0}^2}\frac{d^2\boldsymbol{B}}{dt^2} - \bigtriangledown^2\boldsymbol{B} = 0 \tag{B.28}$$

The general form of these equations can be written as:

$$\frac{1}{{c_0}^2}\frac{d^2\boldsymbol{U}}{dt^2} - \bigtriangledown^2\boldsymbol{U} = 0 \tag{B.29}$$

This becomes important in various areas of optics. The fact that we can represent the electric field and magnetic field as wave equations become important to use other computation tools to analyze electric fields.

B.3 Monochromatic waves

Euler's identity forms a relationship between wave equations and their complex representation and is expressed in Equation B.30.

$$Ae^{j(\omega t+\theta)} = A\cos(\omega t+\theta) + jA\sin(\omega t+\theta) \tag{B.30}$$

In Euler's identity, the cos portion of the wave is the 'real' component while the sin is the 'imaginary' component. When converting between the wave and exponential representations of the wave, it's common to just use the real portion of the wave representation.

A monochromatic wave can be described with the function:

$$u(\mathbf{r},t) = a(\mathbf{r})\cos[2\pi\nu t + \varphi(\mathbf{r})] \tag{B.31}$$

Assuming that this expression represents the real portion of the wave in Euler's identity, we can represent it in the complex form in Equation B.32. Equation B.32 describes the wave completely, containing imaginary and real components, whereas B.31 provides only the real component of the wave representation.

$$U(\mathbf{r},t) = a(\mathbf{r})e^{j\varphi(\mathbf{r})}e^{j2\pi\nu t} \tag{B.32}$$

Equation B.32 can be further modified by noting that the amplitude and the first exponential value are both functions of $\mathbf{r}$. If we let $U(r) = a(\mathbf{r})e^{j\varphi(\mathbf{r})}$, then B.32 becomes:

$$U(\mathbf{r},t) = U(\mathbf{r})e^{j2\pi\nu t} \tag{B.33}$$

The motivation for this becomes apparent when considering the Helmholtz equation.

Helmholtz Equation

Hermann von Helmholtz was a German physician and physicist. Among his contributions to science was the derivation of the Helmholtz equation. This equation is important because it simplifies analysis of physical systems that have spatial and time varying components. It provides a convenient way of separating the spatial and time varying pieces to be individually analyzed. The spatial component of the Helmholtz equation is defined as:

$$\bigtriangledown^2 f = \lambda^2 f \tag{B.34}$$

Here, $\bigtriangledown^2$ is the Laplacian operator. f is the eigenfunction which, in the case for waves, represents amplitude. λ is the eigenvalue which, again for the case of waves, is the wave number.

To get to this equation, we can simplify Equation B.33 to $U(r,t) = U(r)e^{j2\pi\nu t} = A(r)T(t) = AT$.

Given the wave equation from B.29

$$\frac{1}{{c_0}^2}\frac{d^2\boldsymbol{U}}{dt^2} - \bigtriangledown^2\boldsymbol{U} = 0 \tag{B.35}$$

we can substitute the simplified Values for U(r, t) giving:

$$\frac{1}{{c_0}^2}\frac{d^2\boldsymbol{AT}}{dt^2} - \bigtriangledown^2\boldsymbol{AT} = 0 \tag{B.36}$$

and separating to different sides yields:

$$\frac{1}{{c_0}^2}\frac{d^2\boldsymbol{AT}}{dt^2} = \bigtriangledown^2\boldsymbol{AT} \tag{B.37}$$

Moving the spatial components right and temporal left leaves:

$$\frac{1}{{c_0}^2 T}\frac{d^2\boldsymbol{T}}{dt^2} = \frac{\bigtriangledown^2\boldsymbol{A}}{A} \tag{B.38}$$

For Equation B.38 to be true, both sides have to be equal to a constant. If the constant is defined as $-k^2$ and we take the right side of the equation then we get:

$$\frac{\bigtriangledown^2\boldsymbol{A}}{A} = -k^2 \tag{B.39}$$

and finally by multiplying both sides by A and bringing to the same side:

$$\bigtriangledown^2 A + k^2 A = 0 \tag{B.40}$$

which is the Helmholtz equation for the spatial component, or amplitude, of the wave equation.

k is defined as the wave number and is defined as $k = \frac{2\pi\nu}{c} = \frac{\omega}{c}$.

This separation is important as it allows the analysis of the amplitude of a wave without having to deal with the time varying complications from the other elements of the wave equation. A plane wave with a complex amplitude is one of the solutions to the Helmholtz equation.

Paraxial Helmholtz Equation

Within an optical system, an optical axis a central frame of reference that is parallel to rays of light. In real systems, rays are often not exactly parallel, but if they're nearly parallel, they're defined as paraxial rays. This definition holds if the angle from the optical axis are small such that $\cos(\theta) \approx 1$ and $\sin(\theta) \approx \theta$.

A paraxial wave is a wavefront whose normals are paraxial rays. A paraxial wave can be described with

$$U(r) = A(r)e^{-jkz} \tag{B.41}$$

Following a similar procedure with deriving the Helmholtz equation, we can use the paraxial wave with the wave equation to arrive at:

$$\bigtriangledown_T^2 A - j2k\frac{dA}{dz} = 0 \tag{B.42}$$

where $k = \frac{2\pi}{\lambda}$.

The Gaussian beam is one of the useful solutions to the paraxial Helmholtz equation.

Laguerre-Gaussian Beams

A Gaussian beam is a special electromagnetic field whose amplitudes are given by a Gaussian function. For most lasers, this means that the intended output displayed on a surface is a high intensity region in the middle with the intensity falling off as a Gaussian curve at the radius.

The Gaussian beam can be represented by the equation:

$$U(r) = A_0[\frac{W_0}{W(z)}]exp(-\frac{\rho^2}{W^2(z)})exp[-jkz - jk\frac{\rho^2}{2R(z)} + j\zeta(z)] \tag{B.43}$$

Laguerre-Gaussian Beams

A Laguerre-Gaussian beam is another complete set of solutions to the paraxial Helmholtz equation. It is derived by writing the paraxial Helmholtz equation in cylindrical coordinates. As before, we're primarily interested in the complex amplitude. There are two dependencies in the equation, namely l, m, that represent azimuthal and radial indices.

Assuming that the LG wave travels in the z direction, the LG equation is given as:

$$\begin{aligned} U_{l,m}(\rho,\phi,z) = A_{l,m}[\frac{W_0}{W(z)}](\frac{\rho}{W(z)})^l L_m^l(\frac{2\rho^2}{W^2(z)}) exp(-\frac{\rho^2}{W^2(z)}) \\ \times exp[-jkz - jk\frac{\rho^2}{2R(z)} \mp jl\phi + j(l+2m+1)\zeta(z)], \end{aligned} \tag{B.44}$$

where ρ is the amplitude of the wave, ϕ is the angle in the plane perpendicular to the z axis, and z is the distance along the z axis.

Orbital Angular Momentum

The Gaussian and LG beams have the same dependency on ρ and z. One of the notable difference between the two equations, however, is the additional phase factor, $e^{\mp jl\phi}$, present in the LG equation. This factor is proportional to the azimuthal phase angle. When l is zero, the wavefront is composed of planewaves. When $|l| > 0$ the wavefront assumes a helical shape and is referred to as a screw dislocation. The sign on l determines whether the wavefront twists left or right. l also determines the number of helices present. For $l = 2$, for example, there are two intertwined yet distinct helices present.

APPENDIX C

Primer on Optimal Transport

C.1 Overview

Optimal transport is a study of how to most efficiently transport mass from one location to another. This area of study has implications in nearly any area where the matter of interest can be represented as mass. This includes water, material, and light. The idea of optimal transport was originally posed in 1781 by Gaspard Monge (16), who was interested in the optimal way of redistributing mass (soil). The problem remained unresolved until the 1980s and 90s when some breakthroughs in mathematics were made.

A Gaussian beam is monochromatic electromagnetic field whose amplitude profiles take the shape of a Gaussian curve. Gaussian beams are useful, because they are solutions to the wave equation for an electromagnetic field. In following sections we will see that solutions to the wave equation can also be solved using optimal transport. Optimal transport is useful because the field provides a variety of tools for solving a class of problems that can be formulated in a specific way.

Gaussian beams are also useful because they can adopt a helical shape after passing through an SLM or phase plate. This is important for the topic of OAM communications.

Image applications have long suffered from distortions due to spatial and temporal fluctuations caused by physical properties in the atmosphere. The path of optical signals is altered due to changes in the index of refraction, which is caused by temperature, humidity, etc. If we consider an image intensity pattern that has just been formed by combining OAM modes together, we have an interesting and unique light distribution. As this pattern propagates through space, it is susceptible to disruptions. We can think of the pattern as regions of densely packed photons. When turbulence arises in the path of the field, there is an interface that is created from the current position of the photon density distributions to the other side of the turbulent surface. That interface acts as many lenses placed next to each other and can cause displacement of photons of light. As turbulence may not be the same everywhere, there is no guarantee that photons will be equally effected. When there are multiple interfaces that are passed through, these photons can move from one intensity distribution to another within the landscape of the image.

From this perspective one can imagine that at the receiver, due to turbulence or other environmental factors, the original photon density distributions can be significantly altered. In order to model the process that is undergone for an image propagating through space, we need a methodology that is able to encapsulate or capture the changes.

To provide a useful solution to this problem we see a confluence from three primary fields. We have the fundamental description of electromagnetic waves propagating in free space provided by Maxwell's equations. These provide mathematical models of how light moves and can be used to derive useful relationships that can be exploited later on. Second, is the field of optimal transport. Originally optimal transport was used to study how to efficiently and optimally distribute physical material from one location to another. The mathematics behind optimal transport have greatly matured in the past 50-60 years. More recently it has proven very useful in providing a framework for understanding and modeling the movement of light mass from one location to another. The final piece comes from computational tools that allow working with these theories in a reasonably useful time frame.

The end objective is to develop an efficient process that is based in the mathematical properties of electric fields and is justified by the underlying physics of propagation. In the end, a model is created that is able to take a 'clean' image, a 'corrupted' image, and is able to create a transform between the two. The final transform is efficient to compute and is invertible. In this context 'invertible' means once we have the transform, we can get a clean image and operate on it such that we get a corrupted image. We can also operate on a corrupted image and get a clean image on the output.

In the following sections we'll discuss and derive Maxwell's equations C.2, provide an overview of optimal transport C.3, and finally touch on one of the computational tools C.4. These sections aren't meant to provide full derivations, but rather highlight some of the important concepts, key takeaways, and show how they come together to provide a cohesive foundation for work done in this dissertation. Full derivations can be found in the referenced literature.

C.2 Maxwell's Equations

Maxwell's equations can be manipulated in a variety of ways. For this conversation, there is a particular goal in mind. In order to form useful extensions into optimal transport, the equations need to be expressed in a specific format. Optimal transport, as will be discussed in Section C.3, is very good at working with problems that can be expressed in a certain mathematical format. Here, we take the steps to express electric fields in that way.

Maxwell's equations and meanings aren't necessarily obvious at first glance. The idea of an area of a field may not be easily definable so the curl operator ($\bigtriangledown$) is used as a way to reduce a field down to a single point. The principle idea being that the integral of a field through a region (flux) will tell us how much of the EM field has passed through the region. We can divide the region into two parts, take the integral of both, and add the results together. If we continually take smaller and smaller integrals, we wind up with a point source of the field. That point source is the curl.

Maxwell's equations describe fundamental properties of electric fields, magnetic fields and their relationship to each other. The equations for electromagnetic radiation are

$$\bigtriangledown \times \mathbf{E}(x) = i\omega\mu_0\mathbf{H}(x) \tag{C.1}$$

$$\bigtriangledown \times \mathbf{H}(x) = -i\omega\epsilon_0\epsilon(x)\mathbf{E}(x) \tag{C.2}$$

$$\mu_0 \bigtriangledown \cdot \mathbf{H}(x) = 0 \tag{C.3}$$

$$\mu_0 \bigtriangledown \cdot (\epsilon(x)\mathbf{E}(x)) = 0 \tag{C.4}$$

where the electric field intensity vector, $\mathbf{E}(x)$, is in (V/m), ω is the angular frequency,

μ_0 is free-space permeability, $\mathbf{H}(x)$ is a magnetic field intensity vector in (A/m), ϵ_0 is the vacuum dielectric constant, and $\epsilon(x)$ is the relative complex permittivity of the atmosphere. The vector x specifies 3-dimensional space where $x \equiv (x_1, x_2, z)$ and z is the direction that the EM wave propagates. EM radiation is assumed to be monochromatic and its time dependence is governed by angular frequency ω.

If we take the curl of Equation C.1 and substitute in Equations (C.2-C.4) we get the wave equation

$$\bigtriangledown^2 \mathbf{E}(x) + \bigtriangledown(\mathbf{E}(x) \cdot \frac{\bigtriangledown \epsilon(x)}{\epsilon(x)}) + k_0^2 \epsilon(x) \mathbf{E}(x) = 0. \tag{C.5}$$

where $k_0 = \sqrt{\epsilon_0 \mu_0} \omega = 1/\lambda_0$ is the wavenumber and λ_0 is the wavelength.

To arrive at equations in the form needed for optimal transport, the vector electric field needs to be replaced with the scalar electric field

$$\mathbf{E}(x) = \Psi(\vec{x}, z) e^{i k_0 z} \tag{C.6}$$

where $\vec{x} = (x_1, x_2)$ is the plane transverse to the direction of propagation. Substituting Equation C.6 into C.5 yields an equation that is similar to the Schrodinger equation. The Madelung transformation (17) can then be used on the resulting equation to generate the following equations

$$\frac{\partial \rho(\vec{x}, z)}{\partial z} + \bigtriangledown_X \cdot (\rho(\vec{x}, z) v(\vec{x}, z)) = 0 \tag{C.7}$$

$$\frac{\partial \rho(\vec{x}, z)}{\partial z} + (v(\vec{x}, z) \cdot \bigtriangledown_X) v(\vec{x}, z) = 2 \bigtriangledown x\gamma(\eta(\vec{x}, z)) \tag{C.8}$$

$$\frac{\partial \phi(\vec{x}, z)}{\partial z} + \frac{1}{2} (\bigtriangledown_X \phi(\vec{x}, z))^2 = 2\gamma(\eta(\vec{x}, z)) \tag{C.9}$$

Equations C.7, C.8, and C.9 describe electric field propagation through the atmosphere in terms of intensity and phase. z refers to the distance in the direction of propagation where 0 would be the starting point and Z would be the location at the imager. $\rho(\vec{x}, z)$ is the image intensity, $\phi(\vec{x}, z)$ is the phase fluctuations from turbulence, $\bigtriangledown x \cdot$ is the transverse plane phase gradient, v is the velocity of the 'mass' being moved, $\eta(\vec{x}, z)$ is perturbations from turbulence. (As a side note, for the application of OAM images, the starting point would be the 'clean' image and the final location would be the 'corrupted' image.)

The physical interpretation of these equations is very interesting. $\rho(\vec{x}, 0)$ is the intensity at the starting point. $\rho(\vec{x}, Z)$ is the image intensities at the imager. At each progressive location in z, image intensities can be moved according to the transverse phase gradient $\bigtriangledown_X \phi(\vec{x}, z)$.

While computational models exist for working with equations C.5, C.7-C.9 ((18; 19; 20; 21)), they are not practical because of their computational cost, incorrect assumptions, or errors. Consequently we want a more efficient way of working with them.

As previously stated, we're interested in formulating the electric field in a form that optimal transport can work with. As optimal transport was originally developed around the idea of moving mass, Maxwell's equations have been transformed to express fields in terms of mass (intensity), velocity, and location. Equations C.7-C.9 are in the form of continuity and momentum equations from the field of fluid mechanics. This formulation allows the electric field propagation to be analyzed as a transport problem.

C.3 Optimal Transport

Optimal transport was initially a study of how to most efficiently move mass from one distribution to another. The theory of optimal transport was originally studied and developed by Gaspard Monge in 1781 (22) and the principles were later matured with efforts from Leonid Kantorovich in 1948 (23) and Yann Brenier (24). Brenier's contributions lead to connections in partial differential equations, fluid mechanics, geometry, probability theory, and functional analysis. All of the research in optimal transport has resulted in a body of literature and mathematical tools for solving a class of problems that are characterized in a similar way.

Recently, the principles of optimal transport have been applied to image data (25; 26). Optimal transport has also been formally derived for electric fields in (27; 28), and they additionally formulate ties to images being transmitted through a turbulent medium.

As an overview for the following conversation, we'll start with the end in mind. Given a set of clean images and corrupted/distorted images, we want to find a framework that allows us to easily learn a mapping between the two. Benefits of this approach include a model that is invertible, computationally efficient, and requires a single pair of images. Rather than reproducing the derivation, the high level procedure will be discussed, with an inclusion of the primary equations of interest to motivate the topic.

Optimal transport seeks to minimize the amount of energy required to move mass from one location to another. To do this, the kinetic energy required to move ($v(\vec{x}, z)$) a mass ($\rho(\vec{x_0}, z)$) over distance $z = [0, Z]$ and over time interval $t = [0, T]$ can be expressed in terms of *action*. In continuum mechanics this is defined mathematically (29) as

$$\mathcal{A} \equiv Z \int_{\mathbb{R}^{\nvDash}} \int_0^Z \rho(\vec{x_0}, z)|v(\vec{x}, z)|^2 dz d\vec{x}. \tag{C.10}$$

The principle of action minimization has been used for deriving equations of motion for a variety of systems. (30) derived Equations 3.7-3.9 given that the condition that intensity is conserved, which is the case for the application of OAM communications of interest. (28; 31) showed that only one solution exists that minimizes Equation C.10 from a source image intensity distribution, $\rho(\vec{x}, 0)$, and destination image intensity distribution, $\rho(\vec{x}, Z)$.

What is needed to map $\rho(\vec{x}, 0)$ to $\rho(\vec{x}, Z)$ is a modification to Equation C.10. The piece needed to do this is in the form of a Lagrangian flow map, which takes coordinates from time $z = 0$ and maps them to time $z = Z$. Mathematically this is expressed as $\vec{x_z} \equiv f(\vec{x}, z)$. This can be rewritten as

$$det(J_f(\vec{x_0}, z))\rho(\vec{x_z}, z) = \rho(\vec{x_0}, z) \tag{C.11}$$

where $J_f(\vec{x_0}, z)$ is the Jacobian of $f(\vec{x_0}, z)$. Equation C.11 delivers the final piece needed for that transformation. It relates a transformation from an initial image to a final image. (32) and (33) show that if we plug Equation C.11 into C.10 we get

$$d_p(0, Z)^2 = \int_{\mathbb{R}^{\nvDash}} ||f(\vec{x_0}, Z) - \vec{x_0}||^2 \rho(\vec{x_0}, 0) d\vec{x}. \tag{C.12}$$

which produces a unique flow map between the two images.

With the physics based justification of a mapping between two images in place, all that is needed is an efficient algorithm to minimize Equation C.12.

C.4 Radon-Cumulative Distribution Transform

The final piece necessary to using optimal transport for the electric field, is a numerical method that will solve Equation C.12. (25) reviewed a variety of approaches for doing this and then proposed an efficient way to perform the mapping. They introduce the Radon-cumulative distribution transform (R-CDT), which is shown to be highly efficient. Code for performing the R-CDT is provided by (26).

The details behind the R-CDT are beyond the scope of this work. More in-depth information can be found in (26; 34).

The R-CDT is used in a number of different ways in this research. It has been very valuable as an efficient tool to provide the mapping between two images and is based on the physics-based properties of optimal transport.

APPENDIX D

Review of State of the Art CNNs

Research presented in this dissertation engages in a search of the best performing architecture for OAM free-space communications under turbulence, and underwater communications under attenuation. Part of the objective is to see if there is an architecture that provides the most robust performance in both mediums under non-ideal conditions.

Here, a discussion of some of the current state of the art CNNs is provided. They each have unique architectures and those unique characteristics are covered.

D.1 AlexNet

AlexNet(35) won the ImageNet competition in 2012 by a significant margin (15.3% error vs. 26.2% error). They used 8 layers: 5 convolution layers and 3 fully connected. They employed max pooling layers, ReLU activation functions, dropout, and normalization as shown in Figure D.1. This architecture has 62.3 million parameters.

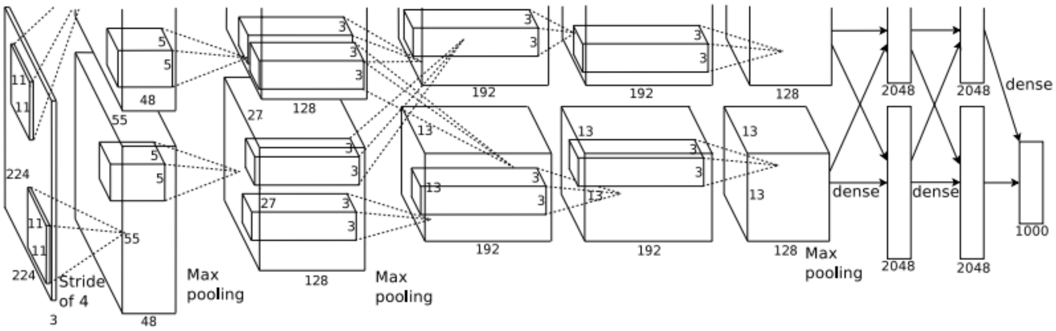


Fig. D.1: AlexNet architecture.

D.2 VGGNet

VGG(36) was the first runner-up for the ImageNet competition in 2014. To date, this was the deepest architecture to win ImageNet at 19 levels. While deeper levels were attempted, it was found that accuracy reached an inflection point and the accuracy of the network dropped. It's believed that this is due to vanishing gradients.

VGG was significant because they used stacked 3x3 kernels. Two layers of 3x3 filters is equivalent to a 5x5 kernel. Three layers of 3x3 filters is equivalent to 7x7, but with a lower parameter count. Consequently, they were able to gain the value of larger filters from stacked 3x3 kernels. This effectively dropped parameter count while also dropping training time. They also reduced the size of the fully connected layer.

A family of architectures (such as ResNet and its variants) are derived from the VGG architecture. Figure D.2 shows how the feature map size reduced through the layers of the network and the number of kernels increased.

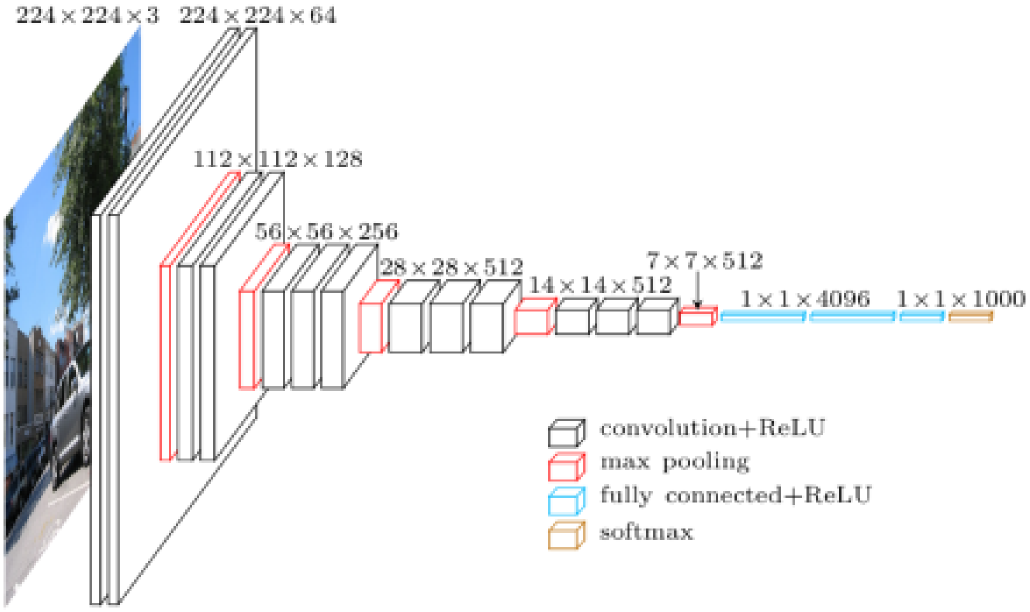


Fig. D.2: VGG architecture.

D.3 Inception/GoogleNet

GoogLeNet/Inception-V1 was the winner for the 2014 ImageNet competition. The Inception model was heavily engineered and employed the idea that kernels of different dimensions could be simultaneously applied to a feature map to extract different information from it.

Figure D.3 shows a diagram of two fundamental Inception Modules. They used max pooling, 1x1, 3x3, and 5x5 kernels in the original implementation (37). In subsequent work they found that the 5x5 kernel could be replaced with stacked 3x3 kernels (38) . This replacement generated the same information, but was cheaper computationally and reduced the number of trainable features.

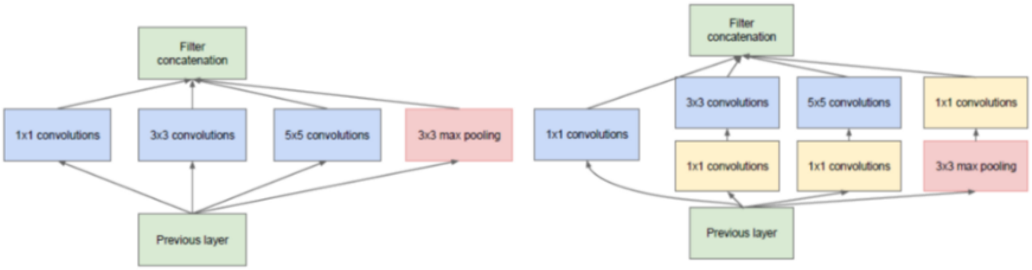


Fig. D.3: Inception Modules.

The Inception model was very effective in the ImageNet competition. The overall architecture is shown in Figure D.4. It has 22 layers, 27 including the pooling layers and uses global average pooling. An interesting contribution from this architecture was the use of auxiliary classifiers (in purple). These helped reduce vanishing gradients during training by applying softmax to two of the inception modules and doing a weighted sum to the real loss.

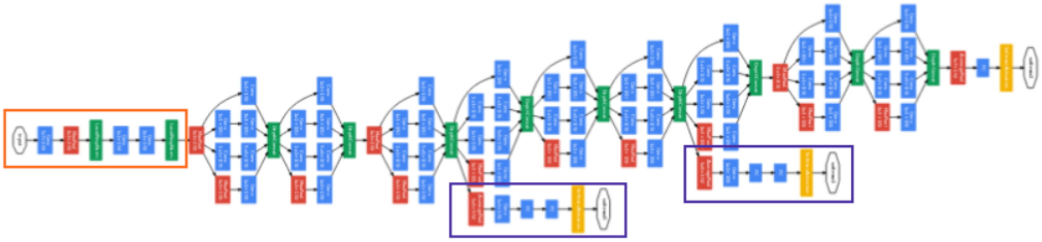


Fig. D.4: Inception architecture.

D.4 ResNet

The ResNet (39) architecture is based on the VGG idea of stacking convolutions. The key contribution of their architecture was the idea of an identity connection. They contend that a key reason VGG failed with deeper networks, was that of accuracy saturation. Due to the depth of the network, the gradients vanished and the network was unable to properly train. By adding identity connections, they were able to propagate information to the deeper layers of the network, and they were able to continue training. They won the ImageNet comptetition in 2015.

By implementing this idea, they were able to create an architecture, identical in composition to the VGGNet, with the sole addition of the identity connection. This new architecture was able to achieve greater accuracy than VGG. Additionally, they were able to create architectures much deeper in which performance continued to improve with the depth. Figure D.5 shows the VGG network, a VGG style network of depth 34, and a 34-layer residual network.

The identity connection is shown in Figure D.6. The identity connection is a simple connection from the input of the block, which is summed with the output. While the identity connection added a trivial amount of additional computation, the results were significant. Identity connections have allowed much deeper, trainable architectures to be created.

D.5 XCeption

The XCeption (40) network was a modification of the Inception network. They took the exact layout of the Inception model, and replaced the 3x3 kernels with depthwise separable convolutions. By using the exact model, with this sole adjustment they were able to achieve a slight improvement in performance, but significant improvement in training speed and reduction in parameter count.

A depthwise convolution performs a convolution on each layer, the results of which are mapped to an output feature map. In a standard convolution, the output feature map will have only one layer for an input of depth M. However, in a depthwise convolution, the output has the same depth as the input.

A pointwise convolution is a 1x1 kernel that is convolved with a feature map set that yields a single layered output for an input set of depth M.

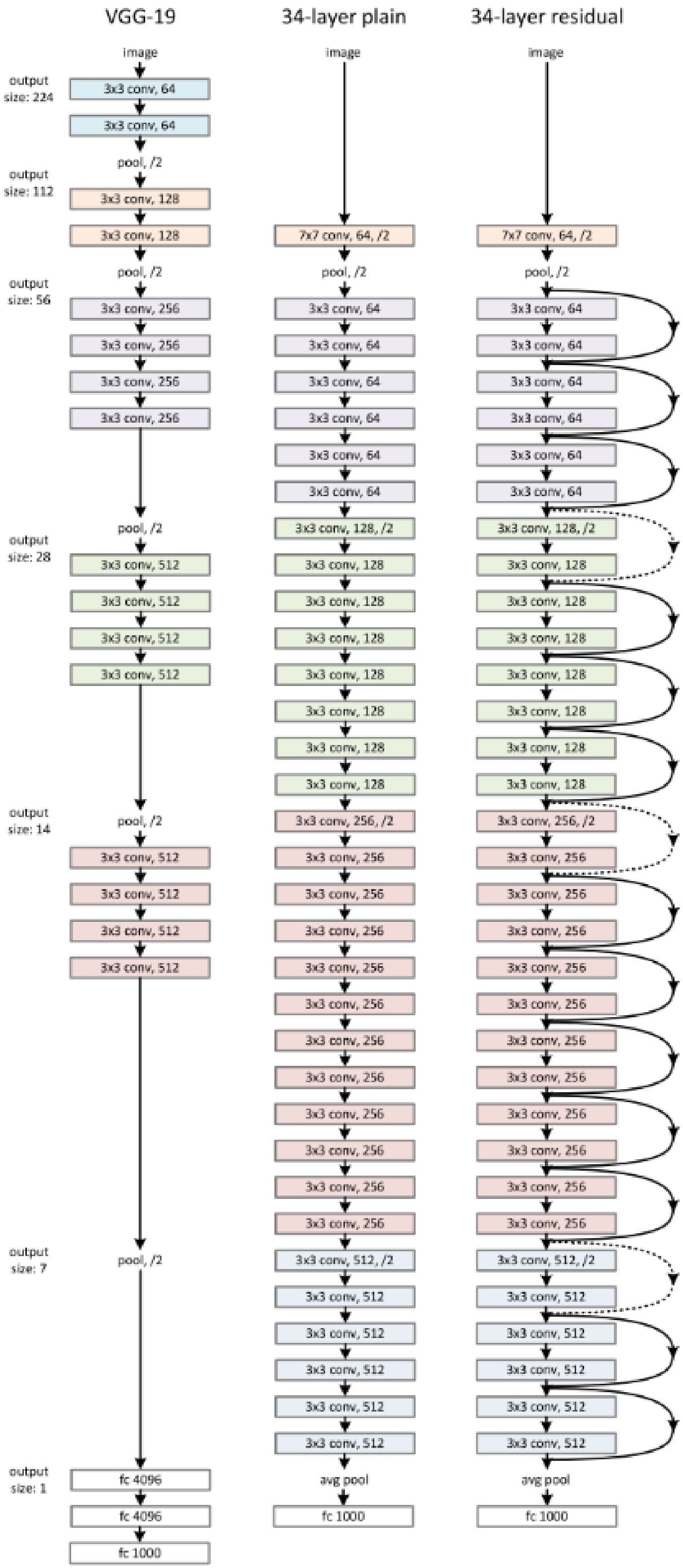


Fig. D.5: ResNet architecture.

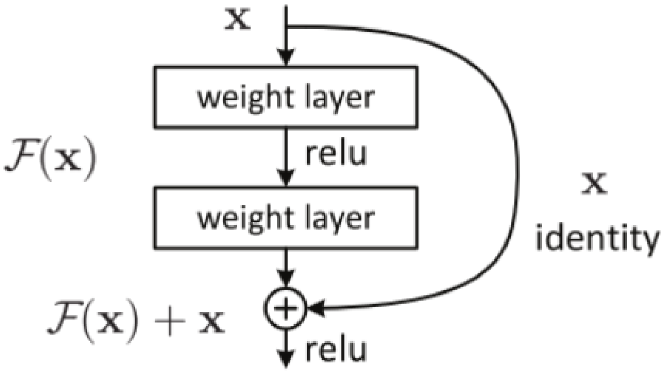


Fig. D.6: Identity connection.

XCeption was able to beat ResNet results in ImageNet by including identity connections in its architecture.

D.6 ResNeXt

ResNeXt (41) is an extension of the ResNet architecture. They postulated that gains could be made through widening the architecture. They introduced the idea of cardinality, where N branches were introduced and each branch contained a small number of kernels as shown in Figure D.7.

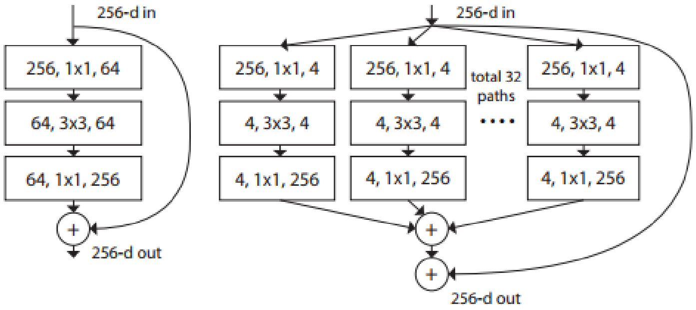


Fig. D.7: ResNeXt block connection.

D.7 DenseNet

DenseNets (42) are inspired by identity connections. The difference from ResNet and its variants, lies in how it propagates information forward. Rather than using an identity connection, it feeds the output from each convolution into the input of all future convolutions. This process is illustrated in Figure D.8.

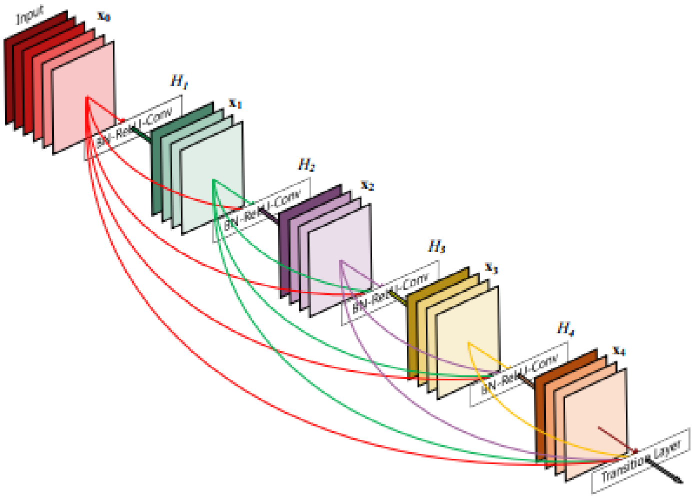


Fig. D.8: Dense block connection.

Consequently, the deeper we go into the convolution series, the higher the number of feature maps. This feature map density helps to alleviate vanishing gradients, encourages feature reuse, and substantially reduces the number of features needed. They showed state of the art performance on several data sets, including ImageNet.

APPENDIX E

Primer on Neural Networks

E.1 History

Early work in neuroscience suggested that neurons fire and propagate their signals on to other neurons. From this idea came the advent of artificial neural networks (ANN). ANNs are patterned loosely off of these initial theories. That is, a neuron receives signals from external sources, they are combined together, and an output signal is generated and passed on.

In 1943 Warren McCulloch and Walter Pitts published a paper on how they thought neurons in the brain might work (43). This was the seminal work that neural networks sprung from. Their approach took N inputs multiplied by N weights, summed them together, and then passed them through a threshold function to produce a binary value. Their model was unable to learn.

In 1958, Frank Rosenblatt published the idea of a perceptron and presented the first model that could learn weights (44). In 1961 Widrow and his students (45) developed Madaline Rule 1 that allowed learning in neural networks.

During these early years neural nets fell out of favor due to hype and an influential book written by Marvin Minsky in 1969 (46). In this book he wrote a compelling argument against the practicality of neural networks due to the prohibitive amount of time required to train them. His work put a chill on research in neural networks for some 12 years.

In 1971 Paul Werdos developed backpropagation for multilayer neural networks (47) in his dissertation. Backpropagation provides a way to calculate the error on the output of the neural network, and then adjust weights in the neural network to minimize that error. This method will find a local minima, but isn't guaranteed to find the global minima.

In 1982 John Hopfield helped reignite some interest in neural networks in academic circles through his paper (48). He effectively created a recurrent neural network. Werdos' work remained a hidden gem until discovered later by Romelhart, Hinton and Williams in

1986 (45). They brought these ideas into the mainstream with an article published in Nature. This article discussed backpropagation with gradient descent as a way to train neurons.

In 1969 LeCunn was able to show neural networks could identify handwritten digits from a handwritten data set provided by the US Postal Service (49). This work used convolutional neural networks. Today's basic feed forward ANNs are composed of perceptrons that have the following components: a set of inputs, weights, a summing function, an activation function, and an output. An example of this is shown in Figure E.1, which has all of these elements. Many perceptrons are often combined together in layers like Figure E.7. The outputs from one layer of perceptrons will form either the inputs to the next layer of perceptrons or produce the outputs from the neural network. A neural network is considered to be 'deep'if it has more more than an input layer and an output layer. Middle layers are often called hidden layers. The perceptron will be discussed in greater detail shortly.

E.2 Basic Building Blocks

E.1 Perceptrons

Fully connected neural networks are built using perceptrons and are generally trained with stochastic gradient descent or one of its variants.

Perceptrons are fundamental building blocks of fully connected neural networks. Figure E.1 shows the structure of a perceptron. On the left are the perceptron inputs. The top input, '1', is a bias term. '$\mathrm{X_1}$' through '$\mathrm{X_m}$' are other inputs. These inputs might be from a data source, such as an image, or from a previous layer. '$\mathrm{w_1}$' through '$\mathrm{w_1}$' are weights applied to each input. These products are then summed together and passed through an activation function. The result is then passed on as the output.

The basic perceptron is able to represent a linear equation of arbitrary length. For example, Equation E.1 represents a line and can be represented with a very basic, one input perceptron with a bias as shown in Figure E.2.

$$y = mx + b \tag{E.1}$$

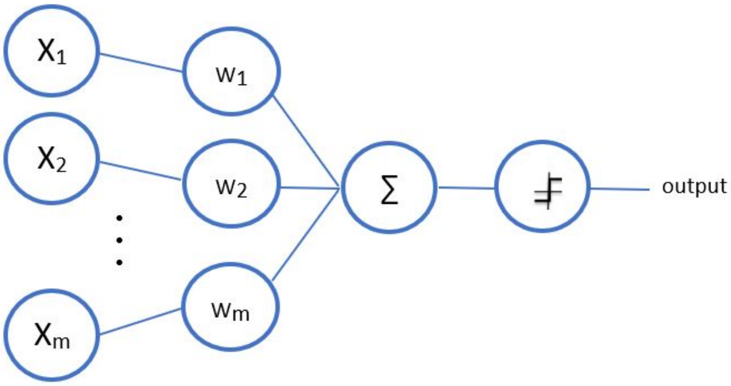


Fig. E.1: Basic perceptron.

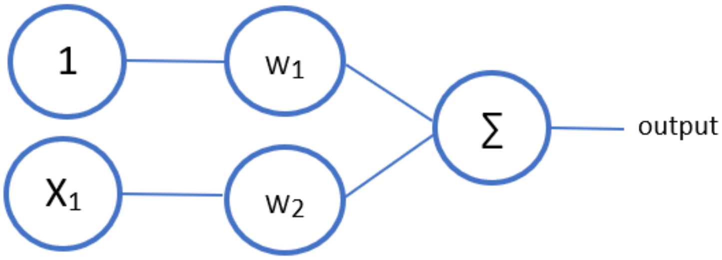


Fig. E.2: Perceptron that can learn a linear equation.

In order to represent more complex equations, a nonlinear element is added to the perceptron. Figure E.3 shows a perceptron with an activation function right before the output. A common activation function is a sigmoid, which has the shape shown in Figure E.4. The sigmoid function maps the linear equation to an output bounded between zero and one.

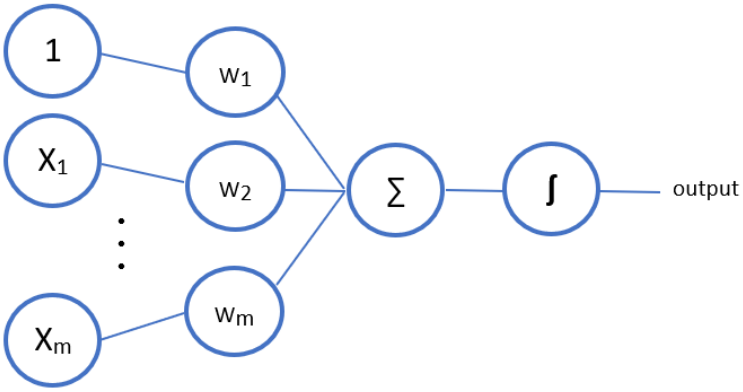


Fig. E.3: Basic perceptron.

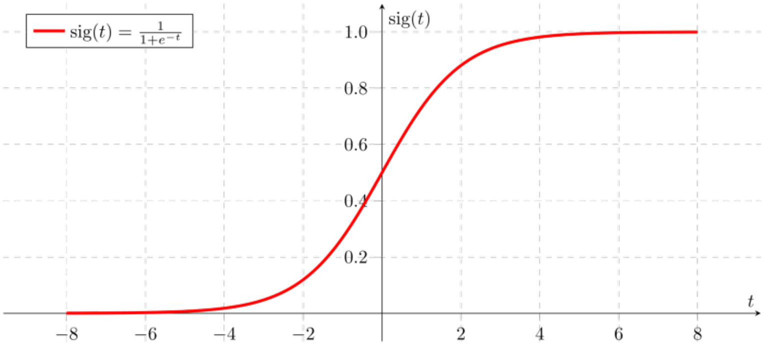


Fig. E.4: Graph and equation of a sigmoid function. (Describe pieces of diagram.)

An important characteristic for any function that serves as an activation function is that it be easily differentiable. This characteristic is critical for gradient descent to work. Additional commonly used activation functions include sigmoid, tanh, and ReLU.

E.2 Gradient Descent

Cost functions enable us to measure the error between expected and actual results. A common cost function is the sum of squared errors (SSE). SSE is shown in Equation E.2. In this equation, 'i' represents the i^{th} sample, where 'output' is a prediction generated by the perceptron and 'target'is the actual value. For example, then network predicted output might be '1.0'when the actual value was supposed to be '1.5'. The output value produced by the network is a function of the weights, ω. Thus the SSE, J, is a function of ω.

$$J(\omega) = \frac{1}{2}\sum_{i}(target^{(\mathrm{i})} - output^{(\mathrm{i})})^2 \tag{E.2}$$

In order to minimize SSE, we can use gradient descent (GD). Gradient descent will allow the weights to be adjusted such that the error is minimized. Figure E.5 shows a convex cost function representing a single weight. The x-axis represents values assigned to the weight, while the y-axis represents the cost function in Equation E.2.

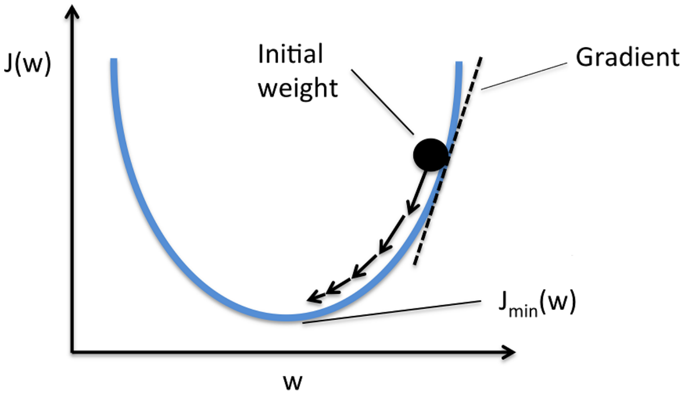


Fig. E.5: Gradient descent.

On the right side of the graph, the gradient for the weight is positive, but we need the weight to drop in value. Consequently, our weight update needs to move in the opposite direction of the gradient. This case also holds if the gradient is on the left side of the graph. Here the gradient is negative, but the weight needs to be increased in order to reach the minimum.

In gradient descent we're interested in adjusting each weight such that the cost function is minimized. To accomplish this, we determine the contribution of each weight to the value of the cost function. This is conveniently accomplished using partial derivatives as shown in Equation E.3. This equation calculates the delta by which to adjust a specific weight.

$$\Delta\omega_{\mathrm{j}} = -\eta \frac{dJ}{d\omega_{\mathrm{j}}} \tag{E.3}$$

$$\omega_{\mathrm{j}} := \omega_{\mathrm{j}} + \Delta\omega_{\mathrm{j}} \tag{E.4}$$

In Equation E.3, ω_{j} is the j^{th} weight in a perceptron. $\Delta\omega_{\mathrm{j}}$ is the amount by which we want to update ω_{j}. η is the learning rate and can take on values between 0.0 and 1.0. $\frac{dJ}{d\omega_{\mathrm{i}}}$ is the partial derivative of the cost function with respect to the weight (ω_{i}).

When the weight is actually updated, Equation E.4 is used. $d\omega_{\mathrm{j}}$ is updated to its previous value plus the delta calculated in Equation E.3.

To solve Equation E.3, we substitute the equation for SSE (E.2) for 'J'. The derivation for the partial derivative of SSE with respect to (ω_{i}) is as follows:

$$\begin{aligned}
\frac{dJ}{d\omega_{\mathrm{j}}} &= \frac{d}{d\omega_{\mathrm{j}}} \frac{1}{2} \sum_{i} (target^{(i)} - output^{(i)})^2 \\
&= \frac{1}{m} \sum_{i}^{m} \frac{d}{d\omega_{\mathrm{j}}} (target^{(i)} - output^{(i)})^2 \\
&= \frac{1}{m} \sum_{i}^{m} 2(target^{(i)} - output^{(i)}) \frac{d}{d\omega_{\mathrm{j}}} (target^{(i)} - output^{(i)}) \\
&= \frac{1}{m} \sum_{i}^{m} (target^{(i)} - output^{(i)}) \frac{d}{d\omega_{\mathrm{j}}} (target^{(i)} - \sum_{i} \omega_{\mathrm{j}} x_{\mathrm{j}}{}^{(i)}) \\
&= \frac{2}{m} \sum_{i}^{m} (target^{(i)} - output^{(i)})
\end{aligned}$$

$$\frac{dJ}{d\omega_{\mathrm{j}}} = \sum_{i}^{m} (target^{(i)} - output^{(i)})(-x_{\mathrm{j}}{}^{(i)}) \tag{E.5}$$

Applying this result back to the learning rate we get:

$$\Delta\omega_{\mathrm{j}} = -\eta\frac{dJ}{d\omega_{\mathrm{j}}} = -\eta\sum_{i}(target^{(i)} - output^{(i)})(-x_{\mathrm{j}}{}^{(i)}) = \eta\sum_{i}(x_{\mathrm{j}}{}^{(i)}) \tag{E.6}$$

This method is called ”batch” gradient descent because it can take a batch of inputs and their corresponding outputs, indexed by i, and update the weights based on results from the entire batch, rather than one sample at a time.

E.3 Learning Rates

An important consideration when performing GD is selecting the learning rate. If the learning rate is too high, we can wind up with wildly swinging values resulting in divergence, as illustrated in Figure E.6. Conversely, if the learning rate is too small it can take a very long time for the weights to converge on their optimal values. Picking good values for the learning rate can reduce training time significantly.

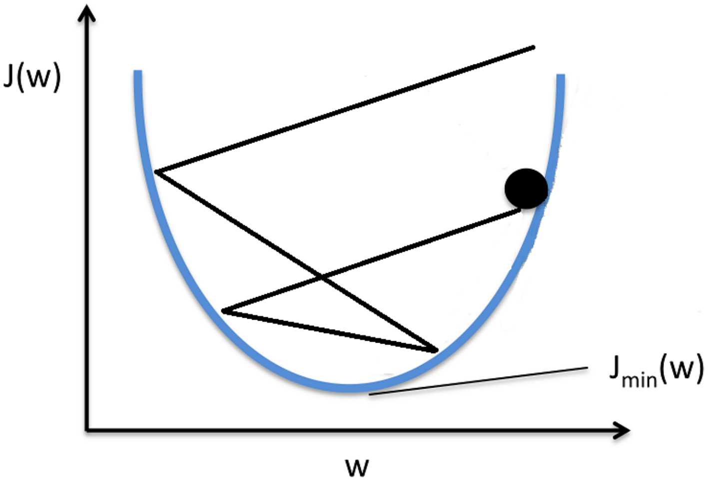


Fig. E.6: Divergence in weights when learning rates are too high.

E.4 Weight Initialization

Another important consideration is how weights are initialized. If weights are initialized to the same value, then when they are updated with GD, they will track each other. A common way to initialize weights is to use random uniform distribution. This breaks symmetry between weights. Simple random uniform initialization works well for shallow networks, but for deep networks, there are better initialization methods.

Xavier initialization, (50), is frequently used to initialize weights of deep networks. This method creates a distribution of weights based for a learnable parameter, based on the number of neurons feeding into the node, plus the number of neurons it feeds into. Equation (E.7) shows this relationship where n_{in} is the number of inputs coming into the neuron and n_{out} is the number of neuron outputs to the next layer. 'Var' is the variance to be applied to the neuron weights, 'W', when initialized.

$$Var(W) = \frac{2}{n_{in} + n_{out}} \tag{E.7}$$

(51) introduce the PReLU and added to the process for initializing weights. These methods were found to improve the classification performance of the networks and surpassed human-level performance.

E.5 Normalization

Another common operation is batch normalization. Batch normalization is the process where inputs to a layer are normalized using Equation E.8, where $\bar{x}$ is calculated from all inputs to neurons in the current layer for a single instance.

$$\hat{x} = \frac{x + mean(\bar{x})}{std(\bar{x})} \tag{E.8}$$

The purpose of normalization is to help with vanishing and exploding gradients (52). Batch normalization is used in various portions of this work.

E.3 Fully Connected, Deep Networks

Perceptrons provide the building block to fully connected networks (FCN). Figure E.7 shows a simple example of a FCN. It has three inputs, 5 perceptrons in the input layer, and two perceptrons in the output layer. The network is considered to be 'fully connected' because each perceptron receives inputs from all sources in the previous layer. Each perceptron also sources its output to all perceptrons in the subsequent layer.

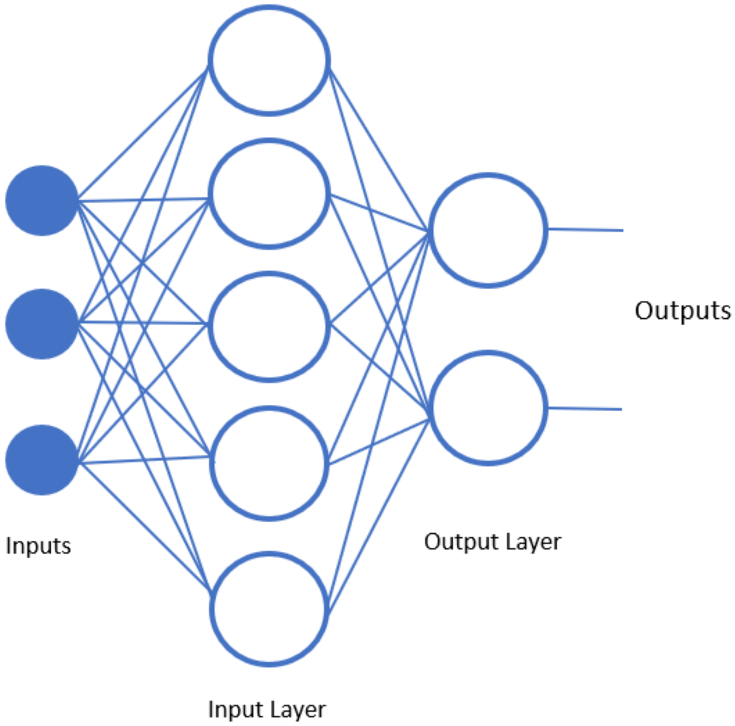


Fig. E.7: Basic fully connected network.

A ”deep” neural network is simply a FCN that has additional layers between the input layer and the output layer. Any layer between the input layer and the output layer is considered to be ‘hidden’. For example Figure E.8 is a deep network with one hidden layer. Generally the input and output dimensions are dictated and fixed by the application. For example the MNIST data set has ten numbers or classes (0-9) and will, therefore, require ten outputs from the network. However, the number of perceptrons in a hidden layer and the number of hidden layers in a FCN may vary according to the design of the network architect.

It is rarely known beforehand what configuration is best for a particular application. In practice, a variety of architectures are tried and the best one is selected for use.

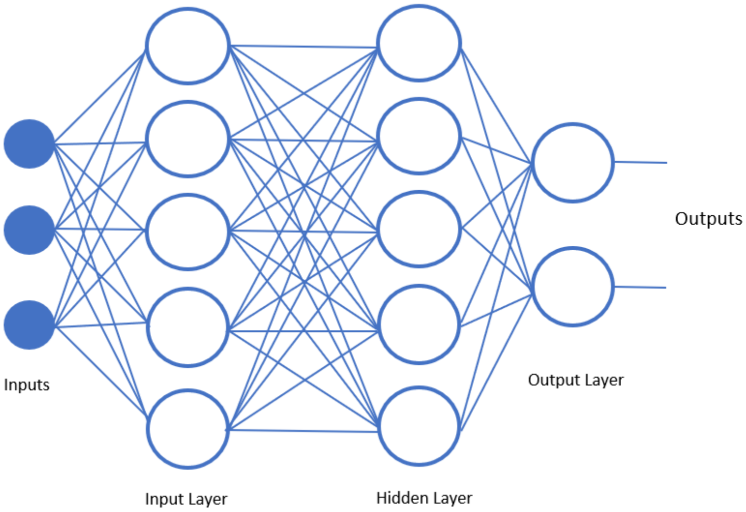
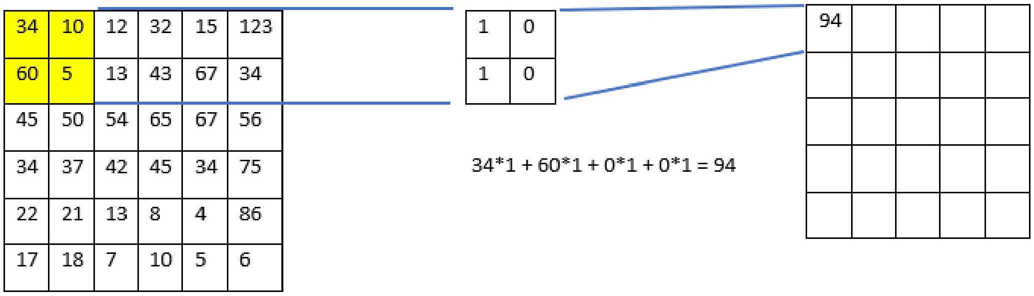


Fig. E.8: Basic fully connected ’deep’ network.

E.4 Convolutional Neural Networks

In the context of neural networks, a convolution is a mathematical operation applied to data presented to the network. Convolutional neural networks are often used with imagery. Given the 6x6 ’image’, 2x2 filter, and 5x5 feature map shown in Figure E.9, we will explore how convolutions work. Filters are also referred to as kernels in the literature.



34	10	12	32	15	123
60	5	13	43	67	34
45	50	54	65	67	56
34	37	42	45	34	75
22	21	13	8	4	86
17	18	7	10	5	6

1	0
1	0

34*1 + 60*1 + 0*1 + 0*1 = 94

94				

Fig. E.9: Basic convolution of filter with image.

To convolve the filter with the image, the filter slides over the top of the image. The overlapping area is called the receptive field. Starting in the upper left-hand corner, the filter is moved left to right and then dropped to the next set of rows. The values of the overlapping cells are multiplied together and then the products are summed. The result is placed in the feature map and the filter is moved to the right.

Generally, there is more than just one layer that is being evaluated. For example, images consist of three channels (red, green, and blue). When a 3x3x3 (width, height, depth) filter is convolved with the image, the output feature map has only one layer. In this example, three 3x3 filters are convolved, one with each layer. This results in 27 numbers that are all added together. The summed value is passed through an activation function and then placed in the new feature map.

The number of columns that the filter slides after each operation is called the horizontal stride. When the filter reaches the right boundary of the image it is moved back to the far left and then drops the number of rows specified by the vertical stride.

As may be noted from Figure E.9, the final matrix in this example will be 3x3 and thus smaller than the original. The reduction in size corresponds to the filter size, stride, and padding. For a stride of one, the horizontal and vertical dimensions will be reduced to one. If the filter size was 3x3, the final image dimensions would be reduced by 2. In Equation E.9 $\mathrm{I_w}$ is the input image width, $\mathrm{F_w}$ is the filter width, $\mathrm{S_w}$ is the horizontal stride, $\mathrm{P_w}$ is the number of padded columns added, and $\mathrm{O_w}$ is the feature map width. The following is true for calculating both the output width and height:

$$O_{\mathrm{w}} = \frac{(I_{\mathrm{w}} - F_{\mathrm{w}} + 2 * P_{\mathrm{w}})}{S_{\mathrm{w}} + 1} \tag{E.9}$$

If it is desired to keep the feature map and input dimensions equal, then padding has to be added to the input matrix. Figure E.10 shows and example of the original image with zero padding added to retain the original dimensions. This 2x2 filter in this figure has a horizontal stride of 1. The calculation for each convolution is shown and the results are placed in the feature map.

E.1 Filters

Each individual filter is tuned to pull out specific features in an image. For example, filters may identify horizontal and vertical features in an image. Generally, deeper layers are able to identify more complex patterns and objects within an image.

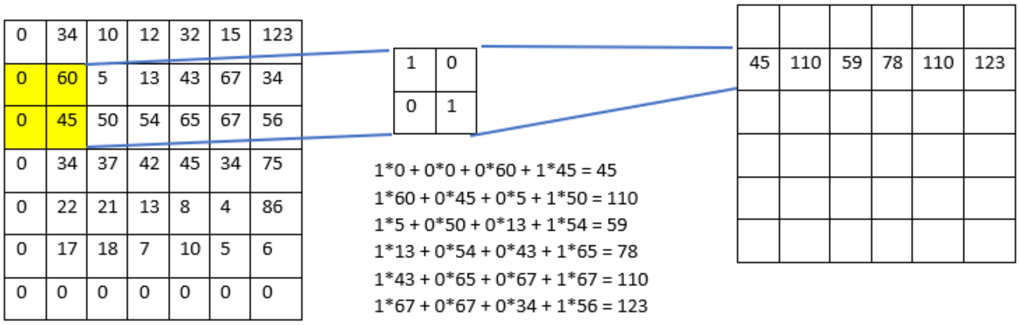


Fig. E.10: Basic convolution of filter with padded image.

Multiple filters can be used in each layer of a convolutional network. In Figure E.10, one filter was applied to an image and generated one feature map. If two filters were used on the image, then two maps would be generated, and so on.

E.2 Pooling

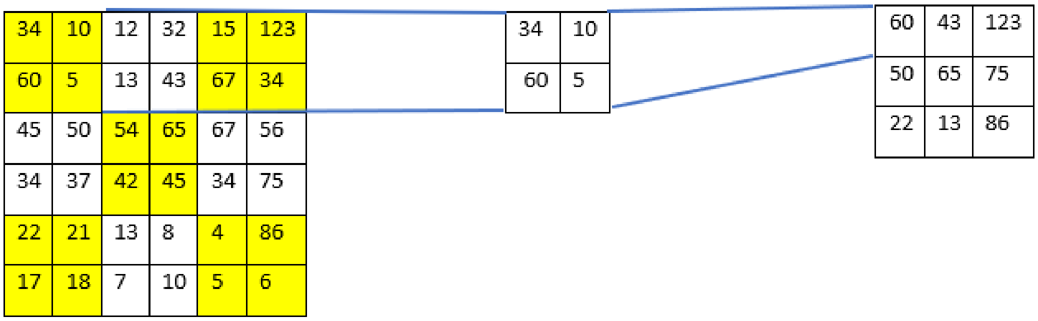
There are a wide variety of ways to modify ConvNet architectures. Two common operations are pooling and normalization.

Pooling is a process that helps reduce feature map dimensions in the ConvNet. Two common types of pooling include max pooling and average pooling. Max pooling takes the maximum value in a defined region, while average pooling averages over all values in a region. Generally, when pooling is applied, a 2x2 or 3x3 shape to slide over a feature map. The pooling shape is moved over the image slice and the maximum or average pixel value of the area within the shape is passed to a receptive field. Figure E.11 provides an example of this operation.

In Figure E.11, max pooling is shown. Max pooling takes the maximum value from each yellow and white blocked section, and writes that value out to the feature map. Different pooling sizes can be selected, but in practice pooling sizes are generally 2x2 or 3x3.

E.3 Fully Connected Layer

On the output of the last convolutional layer, there is usually a fully connected network. More than one fully connected layer may be present. FCLs have the role of identifying



34	10	12	32	15	123
60	5	13	43	67	34
45	50	54	65	67	56
34	37	42	45	34	75
22	21	13	8	4	86
17	18	7	10	5	6

34	10
60	5

60	43	123
50	65	75
22	13	86

Fig. E.11: Basic pooling of a convolution layer.

patterns and predicting results. FCLs can take on any number of layers and any number of perceptrons in each layer.

APPENDIX F

Publication Front Pages

Transport vs. Deep Neural Networks in OAM Underwater Communications

Patrick L. Neary, Nicholas Flann
Department of Computer Science
Utah State University
Logan, Utah, USA
Patrick.Neary@sdl.usu.edu

Abbie T. Watnik, Jonathan M. Nichols, K. Peter Judd, James R. Lindle
Applied Optics Branch
Naval Research Laboratory
Washington, D.C. USA
abbie.watnik@nrl.navy.mil,
jonathan.nichols@nrl.navy.mil, kyle.judd@nrl.navy.mil

Abstract— **Signal attenuation, resulting in low signal to noise ratio (SNR), in underwater optical communications (UWOC) is a problem that degrades classification performance. We develop and contrast several novel ways to create machine learning (ML) and optimal transport-based attenuation models and insert these models in a convolutional neural network (CNN) classification training pipeline. We show that including these ML-based attenuation models in the CNN classifier training, significantly improves classification performance.**

Keywords— ***optimal transport; machine learning models; underwater OAM communications;***

I. INTRODUCTION

OAM communications have great potential in free space or underwater environments to increase communication bandwidth by multiplexing modes together. The idea behind such schemes is to equate the intensity patterns associated with each mode to a particular bit pattern. Messages can then be sent between the source (laser) and receiver (imager) provided that one can successfully recognize the image that was sent.

While strides have been made toward developing a free space OAM communication link, a number of issues remain in the underwater domain. For example, the signal to noise ratio (SNR) between a laser source and an imager can change due to water turbidity.

In this presentation, we discuss a classifier capable of recognizing an attenuated, noisy OAM mode. The classifier is based on prior work [1]. The prior work applied deep learning to a collection of OAM images as a way to learn multiplexed OAM modes in a free space channel corrupted by turbulence (but no attenuation). They used a custom shallow CNN as well as AlexNet for OAM mode classification. Here we extend this work to a signal attenuated in the underwater domain, and use the shallow CNN they presented as a baseline architecture. We are specifically interested in creating machine learning-based models of the attenuated signals, and using those models during training of a classifier, with the intent of improving classifier robustness in a low SNR environment..

II. APPROACH

Four sets of underwater OAM data were collected at varying attenuation lengths (0, 4, 8, and 12). Attenuation length was calculated according to $i=i_0e$^$(-AL)$, where AL is the attenuation length, i is attenuation intensity, and i_0 is the initial intensity. The data sets were divided into training, validation, and test sets (70%, 15%, 15%).

Two machine learning-based models were created, one using a CNN and the other using optimal transport [2] and the Radon-cumulative distribution transform (R-CDT). As the models are trained, they are able to use all information in the AL0-AL12 training sets.

After the ML-based models are trained, they are inserted into the classifier training pipeline. The CNN-based classifier is trained using only data from the AL0 training set.

In evaluating classifier performance, test data from each of the attenuation length sets (AL0-AL12) is presented to the classifier, and their results are aggregated and compared.

III. RESULTS

A CNN classifier is trained without a ML-based attenuation model in the training pipeline. For AL0 it has great results. As the SNR drops with AL4-AL12, accuracy drops sharply. With the CNN ML-model in the loop, results improve significantly. The most compelling results are, however, with the optimal transport/R-CDT-based model.

IV. REFERENCES

[1] Timothy Doster and Abbie T. Watnik. Measuring multiplexed OAM modes with convolutional neural networks. In Lasers Congress 2016. doi: 10.1364/LSC.2016.LTh3B.2.

[2] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde. Transport-based model for turbulence-corrupted imagery. doi: 10.1364/AO.57.004524.

Transport-based pattern recognition vs. deep neural networks in underwater OAM communications

PATRICK L. NEARY[1], JONATHAN M. NICHOLS[1], ABBIE T. WATNIK[1], K. PETER JUDD[1], JAMES R. LINDLE[1], G. K. ROHDE[3], AND NICHOLAS S. FLANN[2]

[1] *Naval Research Laboratory, Washington D.C., 20036*
[2] *Department of Computer Science, Utah State University, Old Main Hill, Logan UT, 84322*
[3] *Dept. of Electrical and Computer Engineering, 415 Lane Road, University of Virginia, Charlottesville, VA 22908*

Compiled May 5, 2020

This paper explores underwater optical communications and compares it to recent findings in free space. The physics justifying optimal transport for use in attenuated water environments is developed. In clear water environments, results are shown to be similar to free space communications. Additionally, some of the complications introduced by signal attenuation are highlighted. The radon cumulative distribution transform (R-CDT) is applied to the OAM patterns and its classification results are compared standard image classification accuracies. To perform classification on the original OAM patterns and their R-CDT, the Nearest Subspace algorithm, shallow convolutional neural network, and deep neural network are used. It is shown that optimal transport and the R-CDT provide better results than CNNs in OAM pattern classification in underwater environments.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

---

1. INTRODUCTION

Free space optical communications utilizing orbital angular momentum has been an area of active interest. A largely unexplored application of these ideas is in the underwater environment.

Digital communication demands are continually increasing. Currently it is estimated that only 58% of the world's population is connected to the Internet [1]. As access continues to be extended to remote areas and usage continues to proliferate, demand for bandwidth will only increase. Difficulties exist in providing access to remote areas, such as costs for laying cables, and dealing with difficult obstacles (in water or land) when laying cables. Other concerns related to communications are the problems of security and dealing with eavesdropping with signals ([2]).

One approach to addressing these issues is to use free space optical (FSO) communications. FSO provides the benefit of bypassing costly cable installations through geographically difficult regions. Another benefit of FSO is the fact that it is difficult to intercept the communication stream without degrading the signal and alerting the system to tampering attempts. However, security is an ongoing area of interest and study with OAM communications [3].

OAM is a property of electromagnetic (EM) waves. As coherent light is created and transmitted, EM radiation propagates in plane waves. In 1992, [4] was able to show, that in addition to having forward momentum, Laguerre-Gaussian (LG) beams also display angular momentum. Assuming that light propagates along a z-axis, they showed that LG beams could be induced to travel in a helical pattern about that axis. The LG equation contains an azimuthal dependency expressed as $exp(-i\ell\phi)$, where ℓ is called the topological charge. When ℓ is 0, the beam propagates in a standard planar wavefront. When $|\ell| > 0$, the EM wave experiences angular momentum and travels in a helically shaped wavefront and the radius of the EM wave increases with ℓ. The sign on ℓ determines whether the EM helix propagates in a left or right handed direction. One of the compelling properties of OAM is that topologically charged modes are orthogonal to each other. Consequently, the beams can be multiplexed together. In ideal transmission environments, the beams can be demultiplexed at the receiver without loss or cross talk of any signals. [5] showed that by multiplexing four OAM modes together, they were able to achieve > 100 Tera-bit/s data rates. This was accomplished over short distances in an ideal environment.

Work continues to move forward in addressing the confounding properties of environmental conditions. For example, turbulence in free space environments causes distortions in the patterns at the receiver in a FSO system. This amounts to cross talk between multiplexed channels. Several approaches have

Optics Communications 474 (2020) 126058

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Machine learning-based signal degradation models for attenuated underwater optical communication OAM beams

Patrick L. Neary [a,*], Abbie T. Watnik [b], K. Peter Judd [b], James R. Lindle [c], Nicholas S. Flann [a]

[a] *Utah State University, Logan, UT 84322, USA*
[b] *Naval Research Laboratory, Washington, DC 20036, USA*
[c] *DCS Corporation, Alexandria, VA 22310, USA*

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Convolutional neural networks
Automatic differentiation
Radon cumulative distribution transform
Attenuation models
Orbital angular momentum
Underwater communications
Physics-based training

ABSTRACT

Signal attenuation in underwater communications is a problem that degrades classification performance. Several novel CNN-based (SMART) models are developed to capture the physics of the attenuation process. One model is built and trained using automatic differentiation and another uses the radon cumulative distribution transform. These models are inserted in the classifier training pipeline. It is shown that including these attenuation models in classifier training significantly improves classification performance when the trained model is tested with environmentally attenuated images. The improved classification accuracy will be important in future OAM underwater optical communication applications.

1. Introduction

Orbital Angular Momentum (OAM) in electromagnetic waves has created a stir in communications research [1,2]. OAM communications have great potential in free space or underwater [3] environments to increase communication bandwidth by multiplexing modes together. The idea behind such schemes is to equate the distinct intensity patterns associated with each mode to a particular bit pattern. Messages can then be sent between the source (laser) and receiver (imager) provided that one can successfully recognize the image that was sent.

Orbital angular momentum beams provide a unique method to increase the capacity of underwater communications using the spatial dimension; for reference on work by others, see [4]. In this paper, an OAM underwater optical communication approach was implemented using 16 spatial modes (including the null set). The analysis provided in this paper focuses on the classification of these modes as an important step to full implementation of an underwater OAM system.

Several significant limiting factors in OAM communications exist, these include both turbulence [5] and signal attenuation [6]. Attenuation is a result of the total absorption and scattering experienced in a medium such as water. The power loss due to attenuation can be described as $P_r = P_0 exp^{-(a+b)z}$, where P_r is received power, P_0 is source power, a is the absorption coefficient, b is the scattering coefficients, and z is the range [7]. Attenuation from particulates is anticipated to occur in any open system as natural events, such as storms, stir up sediment in the water.

While strides have been made toward developing a free space OAM communication link [8,9], a number of issues remain in the underwater domain. For example, the signal to noise ratio (SNR) between a laser source and an imager can change due to water turbidity. SNR in clear water may be initially high, but weather conditions may stir up particulates in the water column and cause it to decrease. Additionally, simply increasing the distance between source and receiver can cause a significant drop in SNR. Consequently it becomes important to investigate classifier performance and robustness in these situations.

A potential issue inherent in any classification system is how well it performs when it is given data outside of its training set. The objective becomes, how to use a single data set and extend it such that the trained classifier continues to perform well when the environment degrades beyond the conditions of the original training set. Also, given that environments can change drastically, what can be done to quickly and easily update a classification algorithm when conditions change beyond what is originally anticipated?

This work, therefore, seeks to develop a classifier capable of recognizing attenuated (noisy) OAM modes. The classifier is based on the prior work of [10] as well as that of [11]. These cited works applied deep learning to a collection of OAM images as a way to

---

[*] Corresponding author.

E-mail addresses: patrick.neary@sdl.usu.edu (P.L. Neary), abbie.watnik@nrl.navy.mil (A.T. Watnik), kyle.judd@nrl.navy.mil (K.P. Judd), james.lindle.ctr@nrl.navy.mil (J.R. Lindle), Nicholas.Flann@usu.edu (N.S. Flann).

https://doi.org/10.1016/j.optcom.2020.126058
Received 31 March 2020; Received in revised form 8 May 2020; Accepted 9 May 2020

CNN Classification Architecture Study for Turbulent Free-Space and Attenuated Underwater Optical OAM Communications

Patrick Neary[1*], **k. Peter Judd**[2], **Abbie T. Watnik**[2], **Ryan R. Lindle**[3], and **Nicholas Flann**[1]

[1] *Utah State University, Logan, UT 84322, USA*
[2] *U.S. Naval Research Laboratory, Washington, DC 20375, USA*
[3] *DCS Corporation, Alexandria, VA 22310, USA*
[*] *Corresponding author: patrick.neary@sdl.usu.edu*

Compiled May 22, 2020

Turbulence and attenuation are signal degrading factors that can severely hinder free-space and underwater OAM optical pattern demultiplexing. A variety of state-of-the-art convolutional neural network architectures are explored to identify which, if any, provide optimal performance under these non-ideal environmental conditions. Hyperparameter searches are performed on the architectures to ensure that near-ideal settings are used for training. Architectures are compared in various scenarios and the best performing, with their settings, are provided. We show that from the current state-of-the-art architectures, DenseNet outperforms all others when memory is not a constraint. When memory footprint is a factor, ShuffleNet is shown to performed the best.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

In 2017, Doster and Watnik explored the use of machine learning (ML) to demultiplex OAM beam patterns for free-space optical communications [1]. Since then, ML techniques have been applied in a variety of ways to improve communication accuracy and bit error rates (BER) in free-space turbulent conditions [2, 3].

In OAM communications, turbulence and attenuation can cause significant degradation of signal integrity and lowering of the signal to noise ratio (SNR) [4, 5]. These disturbances can displace spatial patterns, thereby causing crosstalk, or scatter the signals such that only a portion of the original intensity distribution makes it to the receiver.

One of the unresolved questions from Ref. [1] is with regards to which, if any, of the state-of-the-art convolutional neural network (CNN) architectures performs best for OAM pattern demultiplexing in signal degrading environments. This paper explores turbulent free-space and attenuated underwater OAM optical communications with the state-of-the-art deep convolutional neural networks to answer this question.

Several data sets under varying environmental conditions are used for this effort. In free-space, three sets of data are collected at different simulated turbulence levels. In water, four sets of data are collected at various attenuation levels. All tests are performed on specific combinations of these data sets.

Contributions of this paper include a comparison of recent, state-of-the-art CNN architecture in both turbulent free-space and attenuated underwater OAM communications. Baseline performance, inter-set performance, and parameter count are analyzed. At the end of the analysis, the best performing architectures, along with their parameters, are provided.

2. BACKGROUND AND PRIOR ART

In the following sections OAM communications, hyperparameter tuning, and an overview of some of the current state-of-the-art CNNs are covered.

A. Orbital Angular Momentum

Orbital angular momentum (OAM) in electric fields was discovered by Allen et al. [6]. They found that under certain conditions, the Laguerre-Gauss beam could transition from a standard plane wave propagation to a helical path. Consequently, the Gaussian-shaped distribution frequently exhibited by lasers becomes a doughnut shaped pattern when an OAM mode is adopted. The OAM azimuthal dependency is expressed by $exp(i\ell\phi)$, where ℓ is the topological charge or mode number. When $\ell = 0$, the wavefront is a plane. When $|\ell| > 0$, the wavefront travels in a helical path, where the direction of rotation about the z-axis is controlled by the sign on ℓ. The radial distance from the z-axis

REFERENCES

[1] R. A. Beth, “Mechanical detection and measurement of the angular momentum of light,” *Physical Review*, vol. 50, no. 2, pp. 115–125, 1936. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.50.115

[2] L. Allen, M. Beijersbergen, R. Spreeuw, and J. Woerdman, “Orbital angular momentum of light and transformation of laguerre gaussian laser modes,” *Physical review. A*, vol. 45, pp. 8185–8189, 07 1992.

[3] J. Durnin, J. J. Miceli, and J. H. Eberly, “Diffraction-free beams,” *Physical Review Letters*, vol. 58, no. 15, pp. 1499–1501, 1987. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.58.1499

[4] F. Gori, G. Guattari, and C. Padovani, “Bessel-gauss beams,” *Optics Communications*, vol. 64, no. 6, pp. 491–495, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0030401887902768

[5] A. E. Siegman, “Hermite–gaussian functions of complex argument as optical-beam eigenfunctions,” *JOSA*, vol. 63, no. 9, pp. 1093–1094, 1973. [Online]. Available: https://www.osapublishing.org/josa/abstract.cfm?uri=josa-63-9-1093

[6] M. A. Bandres and J. C. Gutiérrez-Vega, “Ince–gaussian beams,” *Optics Letters*, vol. 29, no. 2, pp. 144–146, 2004. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-29-2-144

[7] J. C. Gutiérrez-Vega, M. D. Iturbe-Castillo, and S. Chávez-Cerda, “Alternative formulation for invariant optical fields: Mathieu beams,” *Optics Letters*, vol. 25, no. 20, pp. 1493–1495, 2000. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-25-20-1493

[8] N. R. Heckenberg, R. McDuff, C. P. Smith, and A. G. White, “Generation of optical phase singularities by computer-generated holograms,” *Optics Letters*, vol. 17, no. 3,

pp. 221–223, 1992. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-17-3-221

[9] M. W. Beijersbergen, R. P. C. Coerwinkel, M. Kristensen, and J. P. Woerdman, "Helical-wavefront laser beams produced with a spiral phaseplate," *Optics Communications*, vol. 112, no. 5, pp. 321–327. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0030401894906386

[10] M. W. Beijersbergen, L. Allen, H. E. L. O. van der Veen, and J. P. Woerdman, "Astigmatic laser mode converters and transfer of orbital angular momentum," *Optics Communications*, vol. 96, no. 1, pp. 123–132, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/003040189390535D

[11] K. Volke-Sepúlveda, A. O. Santillán, and R. R. Boullosa, "Transfer of angular momentum to matter from acoustical vortices in free space," *Physical Review Letters*, vol. 100, no. 2, p. 024302. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.100.024302

[12] G. M. Gibson, E. Toninelli, S. A. R. Horsley, G. C. Spalding, E. Hendry, D. B. Phillips, and M. J. Padgett, "Reversal of orbital angular momentum arising from an extreme doppler shift," *Proceedings of the National Academy of Sciences*, vol. 115, no. 15, pp. 3800–3803, 2018.

[13] L. Wang, W. Park, C. Yang, H. Brüns, D. G. Kam, and C. Schuster, "Wireless communication of radio waves carrying orbital angular momentum (OAM) above an infinite ground plane," *IEEE Transactions on Electromagnetic Compatibility*, pp. 1–8, 2020.

[14] R. Chen, H. Zhou, M. Moretti, X. Wang, and J. Li, "Orbital angular momentum waves: Generation, detection and emerging applications," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[15] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, "Demultiplexing vortex modes in optical communications using transport-based pattern recognition," *Optics Express*, vol. 26, no. 4, pp. 4004–4022, 2018.

[16] W. Gangbo and R. J. McCann, “The geometry of optimal transportation,” *Acta Mathematica*, vol. 177, no. 2, pp. 113–161, 1996. [Online]. Available: https://doi.org/10.1007/BF02392620

[17] P. Vadasz, “Rendering the navier-stokes equations for a compressible fluid into the schrödinger equation for quantum mechanics,” *Fluids*, vol. 1, no. 2, p. 18, 2016. [Online]. Available: https://nau.pure.elsevier.com/en/publications/rendering-the-navier-stokes-equations-for-a-compressible-fluid-in

[18] H. Liu, S. Osher, and R. Tsai, “Multi-valued solution and level set methods in computational high frequency wave propagation,” *Communications in Computational Physics*, vol. 1, no. 5, pp. 765–804, 2006. [Online]. Available: http://global-sci.org/intro/article_detail/cicp/7979.html

[19] J.-D. Benamou, O. Lafitte, I. Solliec, and R. Sentis, “A geometric optics method for high-frequency electromagnetic fields computations near fold caustics—part II. the energy,” *Journal of Computational and Applied Mathematics*, vol. 167, no. 1, pp. 91 – 134, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037704270300880X

[20] S. Jin, P. Markowich, and C. Sparber, “Mathematical and computational methods for semiclassical schrödinger equations*,” *Acta Numerica*, vol. 20, pp. 121–209. [Online]. Available: https://www.cambridge.org/core/journals/acta-numerica/article/mathematical-and-computational-methods-for-semiclassical-schrodinger-equations/F2E8B7C4F924DD74680D6632D9725F41

[21] J. Bec and K. Khanin, “Burgers turbulence,” *Physics Reports*, vol. 447, no. 1, pp. 1 – 66, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370157307001457

[22] Gaspard Monge, *Memoire Sur La theorie Des deblais Et Des Rembla.* De L’Imprimerie Roy, 1781.

[23] L. V. Kantorovich, "On a problem of monge," *Journal of Mathematical Sciences*, vol. 133, no. 4, pp. 1383–1383, 2006. [Online]. Available: https://doi.org/10.1007/s10958-006-0050-9

[24] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160440402

[25] S. Kolouri, S. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Transport-based analysis, modeling, and learning from signal and data distributions," *CoRR*, vol. abs/1609.04767, 2016. [Online]. Available: http://arxiv.org/abs/1609.04767

[26] S. Kolouri, A. B. Tosun, J. A. Ozolek, and G. K. Rohde, "A continuous linear optimal transport approach for pattern analysis in image datasets," *Pattern Recognition*, vol. 51, pp. 453–462, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320315003507

[27] J. M. Nichols, A. T. Watnik, T. Doster, S. Park, A. Kanaev, L. Cattell, and G. K. Rohde, "An optimal transport model for imaging in atmospheric turbulence," 2017, arXiv:1705.01050.

[28] J. M. Nichols, T. H. Emerson, L. Cattell, S. Park, A. Kanaev, F. Bucholtz, A. Watnik, T. Doster, and G. K. Rohde, "Transport-based model for turbulence-corrupted imagery," *Applied Optics*, vol. 57, no. 16, pp. 4524–4536, 2018. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-57-16-4524

[29] Leonard Meirovitch, *Principles and techniques of vibrations.* Upper Saddle River, N.J. : Prentice Hall, 1997.

[30] T. E. Gureyev, A. Roberts, and K. A. Nugent, "Phase retrieval with the transport-of-intensity equation: matrix solution with use of zernike polynomials," *Journal of the Optical Society of America A*, vol. 12, no. 9, pp. 1932–1941, Sep 1995. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-12-9-1932

[31] Cédric Villani, *Optimal Transport: Old and New.* Springer Science Business Media, 2008.

[32] J.-D. Benamou and Y. Brenier, “A computational fluid mechanics solution to the monge-kantorovich mass transfer problem,” *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393. [Online]. Available: https://doi.org/10.1007/s002110050002

[33] J.-G. Liu, R. L. Pego, and D. Slepčev, “Least action principles for incompressible flows and geodesics between shapes,” 2016, arXiv:1604.03387.

[34] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde, “The cumulative distribution transform and linear pattern classification,” *Applied and Computational Harmonic Analysis*, vol. 45, no. 3, pp. 616–641, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520317300076

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[36] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556 [cs]*, Sep. 2014, arXiv: 1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” 2014. [Online]. Available: https://arxiv.org/abs/1409.4842v1

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[40] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016. [Online]. Available: https://arxiv.org/abs/1610.02357v3

[41] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987–5995.

[42] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[43] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943. [Online]. Available: https://doi.org/10.1007/BF02478259

[44] F. Rosenblatt, "The perceptron, a perceiving and recognizing automaton project para," *Cornell Aeronautical Laboratory*, Jan. 1957.

[45] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986. [Online]. Available: https://www.nature.com/articles/323533a0

[46] M. L. Minsky and S. A. Papert, *Perceptrons: Expanded Edition.* Cambridge, MA, USA: MIT Press, 1988.

[47] P. J. Werbos, "Beyond regression : new tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Jan. 1974.

[48] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982.

[49] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[50] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Mar. 2010, pp. 249–256. [Online]. Available: http://proceedings.mlr.press/v9/glorot10a.html

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *arXiv:1502.01852 [cs]*, Feb. 2015, arXiv: 1502.01852. [Online]. Available: http://arxiv.org/abs/1502.01852

[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv:1502.03167 [cs]*, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

CURRICULUM VITAE

Patrick L. Neary

Published Journal Articles

- Transport-based pattern recognition vs. deep neural networks in underwater OAM communications, Patrick L. Neary, Jonathan M. Nichols, Abbie T. Watnik, K. Peter Judd, Gustavo K. Rohde, James R. Lindle, Nicholas S. Flann, *Optics Express* , *Submitted.*
- Machine learning-based signal degradation models for attenuated underwater optical communication OAM beams, Patrick L. Neary, Abbie T. Watnik, K. Peter Judd, James R. Lindle, Nicholas S. Flann, *Optics Communications*, *Accepted.*
- Optimal Architectures for OAM Based Underwater Optical Communications in Attenuated, Turbulent Environments, Patrick Neary, Jonathan M. Nichols, Abbie T. Watnik, K. Peter Judd, James Lindle, Nicholas Flann, *Optics Communications*, *Submitted.*

Published Conference Papers

- In-situ Detection of organics and biomolecules via native fluorescence, Heather Smith, Chris Mckay, Andrew Duncan, Anne Anderson, Ron Sims, Patrick Neary, Christopher Lloyd, in *42nd COSPAR Scientific Assembly* , 2017.
- Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning, Patrick Neary, in *Proc. IEEE Int. Conf. on Cognitive Computing (IEEE ICCC)*, 2018.
- Plug and Play Deep Convolutional Neural Networks, Patrick Neary, Vicki Allan, in *Proc. 11th International Conference on Agents and Artificial Intelligence (ICAART)*, 2019.

- Optical communication in maritime environments using orbital angular momentum, K. Peter Judd, Patrick Mahan, Nicholas Makrakis, J. R. Lindle, Abbie T. Watnik, Wade Freeman, Robert Sagusti, Patrick Neary, Charles Payne, Carlos Font, *Proc. SPIE Defense and Commercial Sensing* , 2020.

Invited Talks

- Transport vs. Deep Neural Networks in OAM Underwater Communications, Patrick L. Neary, Abbie T. Watnik, K. Peter Judd, James R. Lindle, *Photonics North* , May 2020.
- THE ASK (ASTROBIOLOGY SEARCH KIT) INSTRUMENT, Heather Smith, Andrew Duncan, Christopher Lloyd, Les Merrill, Patrick L. Neary, *49th Lunar and Planetary Science Conference* , 2018.