

Lei Huang

# Normalization Techniques in Deep Learning

---

# Synthesis Lectures on Computer Vision

## Series Editors

Gerard Medioni, University of Southern California, Los Angeles, CA, USA

Sven Dickinson, Department of Computer Science, University of Toronto, Toronto, ON,  
Canada

This series publishes on topics pertaining to computer vision and pattern recognition. The scope follows the purview of premier computer science conferences, and includes the science of scene reconstruction, event detection, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, and image restoration. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems, such as those in self-driving cars/navigation systems, medical image analysis, and industrial robots.

---

Lei Huang

# Normalization Techniques in Deep Learning

Lei Huang  
Beihang University  
Beijing, China

ISSN 2153-1056 ISSN 2153-1064 (electronic)  
Synthesis Lectures on Computer Vision  
ISBN 978-3-031-14594-0 ISBN 978-3-031-14595-7 (eBook)  
<https://doi.org/10.1007/978-3-031-14595-7>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG  
2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

---

## Preface

I focused my research on normalization techniques in deep learning since 2015, when the milestone technique—batch normalization (BN)—was published. I witness the research progresses of normalization techniques, including the analyses for understanding the mechanism behind the design of corresponding algorithms and the application for particular tasks. Despite the abundance and ever more important roles of normalization techniques, there is an absence of a unifying lens with which to describe, compare, and analyze them. In addition, our understanding of theoretical foundations of these methods for their success remains elusive.

This book provides a research landscape for normalization techniques, covering methods, analyses, and applications. It can provide valuable guidelines for selecting normalization techniques to use in training DNNs. With the help of these guidelines, it will be possible for students/researchers to design new normalization methods tailored to specific tasks or improve the trade-off between efficiency and performance. As the key components in DNNs, normalization techniques are links that connect the theory and application of deep learning. We thus believe that these techniques will continue to have a profound impact on the rapidly growing field of deep learning, and we hope that this book will aid researchers in building a comprehensive landscape for their implementation.

This book is based on our survey paper [1], but with significant extents and updates, including but not limited to the details of techniques and the recent research progresses. The target audiences of this book are graduate students, researchers, and practitioners who have been working on development of novel deep learning algorithms and/or their application to solve practical problems in computer vision and machine learning tasks. For intermediate and advanced level researchers, this book presents theoretical analysis of normalization methods and the mathematical tools used to develop new normalization methods.

Beijing, China  
February 2022

Lei Huang

## Reference

1. Huang, L., J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao (2020). Normalization techniques in training DNNs: Methodology, analysis and application. *CoRR abs/2009.12836*.

---

## Acknowledgements

I would like to express my sincere thanks to all those who worked with me on this book. Some parts of this book are based on our previous published or pre-printed papers [1–9], and I am grateful to my co-authors in the work of normalization technique in deep learning: Xianglong Liu, Adams Wei Yu, Bo Li, Jia Deng, Dawei Yang, Bo Lang, Dacheng Tao, Yi Zhou, Jie Qin, Fan Zhu, Li Liu, Ling Shao, Lei Zhao, Diwen Wan, Yang Liu, Zehuan Yuan. In addition, this work is supported by the National Natural Science Foundation of China (No.62106012). I would also thank all the editors, reviewers, and staff who helped with the publication of the book. Finally, I thank my family for their wholehearted support.

Beijing, China  
February 2022

Lei Huang

---

## References

1. Huang, L., X. Liu, Y. Liu, B. Lang, and D. Tao (2017). Centered weight normalization in accelerating training of deep neural networks. In *ICCV*.
2. Huang, L., X. Liu, B. Lang, and B. Li (2017). Projection based weight normalization for deep neural networks. *arXiv preprint* arXiv:1710.02338.
3. Huang, L., D. Yang, B. Lang, and J. Deng (2018). Decorrelated batch normalization. In *CVPR*.
4. Huang, L., X. Liu, B. Lang, A. W. Yu, Y. Wang, and B. Li (2018). Orthogonal weight normalization: Solution to optimization over multiple dependent stiefel manifolds in deep neural networks. In *AAAI*.
5. Huang, L., Y. Zhou, F. Zhu, L. Liu, and L. Shao (2019). Iterative normalization: Beyond standardization towards efficient whitening. In *CVPR*.
6. Huang, L., L. Zhao, Y. Zhou, F. Zhu, L. Liu, and L. Shao (2020). An investigation into the stochasticity of batch whitening. In *CVPR*.
7. Huang, L., L. Liu, F. Zhu, D. Wan, Z. Yuan, B. Li, and L. Shao (2020). Controllable orthogonalization in training DNNs. In *CVPR*.



8. Huang, L., J. Qin, L. Liu, F. Zhu, and L. Shao (2020). Layer-wise conditioning analysis in exploring the learning dynamics of DNNs. In *ECCV*.
9. Huang, L., Y. Zhou, L. Liu, F. Zhu, and L. Shao (2021). Group whitening: Balancing learning efficiency and representational capacity. In *CVPR*.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Denotations and Definitions	3
1.1.1	Optimization Objective	4
1.1.2	Neural Networks	4
1.1.3	Training DNNs	5
1.1.4	Normalization	6
	References	8
<b>2</b>	<b>Motivation and Overview of Normalization in DNNs</b>	<b>11</b>
2.1	Theory of Normalizing Input	11
2.2	Towards Normalizing Activations	14
2.2.1	Proximal Back-Propagation Framework	14
2.2.2	K-FAC Approximation	15
2.2.3	Highlights of Motivation	16
	References	17
<b>3</b>	<b>A General View of Normalizing Activations</b>	<b>19</b>
3.1	Normalizing Activations by Population Statistics	19
3.2	Local Statistics in a Sample	20
3.3	Batch Normalization	21
	References	24
<b>4</b>	<b>A Framework for Normalizing Activations as Functions</b>	<b>27</b>
4.1	Normalization Area Partitioning	29
4.2	Normalization Operation	30
4.2.1	Beyond Standardization Towards Whitening	30
4.2.2	Variations of Standardization	34
4.2.3	Reduced Standardization	35
4.3	Normalization Representation Recovery	36
	References	39

<b>5</b>	<b>Multi-mode and Combinational Normalization</b>	43
5.1	Multiple Modes	43
5.2	Combination	44
	References	46
<b>6</b>	<b>BN for More Robust Estimation</b>	49
6.1	Normalization as Functions Combining Population Statistics	49
6.2	Robust Inference Methods for BN	50
	References	51
<b>7</b>	<b>Normalizing Weights</b>	53
7.1	Constraints on Weights	54
7.2	Training with Constraints	56
	References	59
<b>8</b>	<b>Normalizing Gradients</b>	61
	References	63
<b>9</b>	<b>Analysis of Normalization</b>	65
9.1	Scale Invariance in Stabilizing Training	66
9.1.1	Auto-Tuning on Learning Rate	70
9.2	Improved Conditioning in Optimization	70
9.3	Stochasticity for Generalization	72
9.3.1	Theoretical Model for Stochasticity	73
9.3.2	Empirical Analyses for Stochasticity	75
9.4	Effects on Representation	79
9.4.1	Constraint on Feature Representation	80
9.4.2	Effect on Representational Capacity of Model	82
	References	83
<b>10</b>	<b>Normalization in Task-Specific Applications</b>	87
10.1	Domain Adaptation	88
10.1.1	Domain Generalization	90
10.1.2	Robust Deep Learning Under Covariate Shift	90
10.1.3	Learning Universal Representations	92
10.2	Image Style Transfer	92
10.2.1	Image Translation	93
10.3	Training GANs	94
10.4	Efficient Deep Models	95
	References	96
<b>11</b>	<b>Summary and Discussion</b>	101
	References	102
	<b>Appendix</b>	105