



	dkozlov Update README.md	27a9563 · 25 days ago	140 commits
	LICENSE	add LICENSE	13 months ago
	README.md	Update README.md	25 days ago

README.md

Awesome Knowledge Distillation

Papers

- Neural Network Ensembles, L.K. Hansen, P. Salamon, 1990
- Neural Network Ensembles, Cross Validation, and Active Learning, Andres Krogh, Jesper Vedelsby, 1995
- Combining labeled and unlabeled data with co-training, A. Blum, T. Mitchell, 1998
- Ensemble Methods in Machine Learning, Thomas G. Dietterich, 2000
- Model Compression, Rich Caruana, 2006
- Dark knowledge, Geoffrey Hinton, Oriol Vinyals, Jeff Dean, 2014
- Learning with Pseudo-Ensembles, Philip Bachman, Ouais Alsharif, Doina Precup, 2014
- Distilling the Knowledge in a Neural Network, Geoffrey Hinton, Oriol Vinyals, Jeff Dean, 2015
- Cross Modal Distillation for Supervision Transfer, Saurabh Gupta, Judy Hoffman, Jitendra Malik, 2015
- Heterogeneous Knowledge Transfer in Video Emotion Recognition, Attribution and Summarization, Baohuan Xu, Yanwei Fu, Yu-Gang Jiang, Boyang Li, Leonid Sigal, 2015
- Distilling Model Knowledge, George Papamakarios, 2015
- Unifying distillation and privileged information, David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, Vladimir Vapnik, 2015
- Learning Using Privileged Information: Similarity Control and Knowledge Transfer, Vladimir Vapnik, Rauf Izmailov, 2015
- Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks, Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, Ananthram Swami, 2016
- Do deep convolutional nets really need to be deep and convolutional?, Gregor Urban, Krzysztof J. Geras, Samira Ebrahimi Kahou, Ozlem Aslan, Shengjie Wang, Rich Caruana, Abdelrahman Mohamed, Matthai Philipose, Matt Richardson, 2016
- Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer, Sergey Zagoruyko, Nikos Komodakis, 2016
- FitNets: Hints for Thin Deep Nets, Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, Yoshua Bengio, 2015
- Deep Model Compression: Distilling Knowledge from Noisy Teachers, Bharat Bhushan Sau, Vineeth N. Balasubramanian, 2016
- Knowledge Distillation for Small-footprint Highway Networks, Liang Lu, Michelle Guo, Steve Renals, 2016
- Sequence-Level Knowledge Distillation, deeplearning-papernotes, Yoon Kim, Alexander M. Rush, 2016
- MobileID: Face Model Compression by Distilling Knowledge from Neurons, Ping Luo, Zhenyao Zhu, Ziwei Liu, Xiaogang Wang and Xiaoou Tang, 2016
- Recurrent Neural Network Training with Dark Knowledge Transfer, Zhiyuan Tang, Dong Wang, Zhiyong Zhang, 2016
- Adapting Models to Signal Degradation using Distillation, Jong-Chyi Su, Subhransu Maji,2016
- Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, Antti Tarvainen, Harri Valpola, 2017
- Data-Free Knowledge Distillation For Deep Neural Networks, Raphael Gontijo Lopes, Stefano Fenu, 2017
- Like What You Like: Knowledge Distill via Neuron Selectivity Transfer, Zehao Huang, Naiyan Wang, 2017
- Learning Loss for Knowledge Distillation with Conditional Adversarial Networks, Zheng Xu, Yen-Chang Hsu, Jiawei Huang, 2017
- DarkRank: Accelerating Deep Metric Learning via Cross Sample Similarities Transfer, Yuntao Chen, Naiyan Wang, Zhaoxiang Zhang, 2017
- Knowledge Projection for Deep Neural Networks, Zhi Zhang, Guanghan Ning, Zhihai He, 2017
- Moonshine: Distilling with Cheap Convolutions, Elliot J. Crowley, Gavin Gray, Amos Storkey, 2017
- Local Affine Approximators for Improving Knowledge Transfer, Suraj Srinivas and Francois Fleuret, 2017
- Best of Both Worlds: Transferring Knowledge from Discriminative Learning to a Generative Visual Dialog Model, Jiasen Lu1, Anitha Kannan, Jianwei Yang, Devi Parikh, Dhruv Batra 2017
- Learning Efficient Object Detection Models with Knowledge Distillation, Guobin Chen, Wengun Choi, Xiang Yu, Tony Han, Manmohan Chandraker, 2017
- Model Distillation with Knowledge Transfer from Face Classification to Alignment and Verification, Chong Wang, Xipeng Lan and Yangang Zhang, 2017
- Learning Transferable Architectures for Scalable Image Recognition, Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le, 2017
- Revisiting knowledge transfer for training object class detectors, Jasper Uijlings, Stefan Popov, Vittorio Ferrari, 2017
- A Gift from Knowledge Distillation: Fast Optimization, Network Minimization and Transfer Learning, Junho Yim, Donggyu Joo, Jihoon Bae, Junmo Kim, 2017
- Rocket Launching: A Universal and Efficient Framework for Training Well-performing Light Net, Zihao Liu, Qi Liu, Tao Liu, Yanzhi Wang, Wujie Wen, 2017
- Data Distillation: Towards Omni-Supervised Learning, Ilija Radosavovic, Piotr Dollár, Ross Girshick, Georgia Gkioxari, Kaiming He, 2017
- Parallel WaveNet:Fast High-Fidelity Speech Synthesis, Aaron van den Oord, Yazhe Li, Igor Babuschkin, Karen Simonyan, Oriol Vinyals, Koray Kavukcuoglu, 2017
- Learning from Noisy Labels with Distillation, Yuncheng Li, Jianchao Yang, Yale Song, Liangliang Cao, Jiebo Luo, Li-Jia Li, 2017
- Deep Mutual Learning, Ying Zhang, Tao Xiang, Timothy M. Hospedales, Huchuan Lu, 2017
- Interpreting Deep Classifiers by Visual Distillation of Dark Knowledge, Kai Xu, Dae Hoon Park, Chang Yi, Charles Sutton, 2018
- Efficient Neural Architecture Search via Parameters Sharing, Hieu Pham, Melody Y. Guan, Barret Zoph, Quoc V. Le, Jeff Dean, 2018
- Transparent Model Distillation, Sarah Tan, Rich Caruana, Giles Hooker, Albert Gordo, 2018
- Defensive Collaborative Multi-task Training - Defending against Adversarial Attack towards Deep Neural Networks, Derek Wang, Chaoran Li, Sheng Wen, Yang Xiang, Wanlei Zhou, Surya Nepal, 2018
- Deep Co-Training for Semi-Supervised Image Recognition, Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, Alan Yuille, 2018
- Feature Distillation: DNN-Oriented JPEG Compression Against Adversarial Examples, Zihao Liu, Qi Liu, Tao Liu, Yanzhi Wang, Wujie Wen, 2018
- Multimodal Recurrent Neural Networks with Information Transfer Layers for Indoor Scene Labeling, Abrar H. Abdulnabi, Bing Shuai, Zhen Zuo, Lap-Pui Chau, Gang Wang, 2018
- Born Again Neural Networks, Tommaso Furlanello, Zachary C. Lipton, Michael Tschannen, Laurent Itti, Anima Anandkumar, 2018
- YASENN: Explaining Neural Networks via Partitioning Activation Sequences, Yaroslav Zharov, Denis Korzhnikov, Pavel Shvechikov, Alexander Tuzhlin, 2018
- Knowledge Distillation with Adversarial Samples Supporting Decision Boundary, Byeongho Heo, Minsik Lee, Sangdoo Yun, Jin Young Choi, 2018
- Knowledge Transfer via Distillation of Activation Boundaries Formed by Hidden Neurons, Byeongho Heo, Minsik Lee, Sangdoo Yun, Jin Young Choi, 2018
- Self-supervised knowledge distillation using singular value decomposition, Seung Hyun Lee, Dae Ha Kim, Byung Cheol Song, 2018
- Multi-Label Image Classification via Knowledge Distillation from Weakly-Supervised Detection, Yongcheng Liu, Lu Sheng, Jing Shao, Junjie Yan, Shiming Xiang, Chunhong Pan, 2018
- Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks, Yuenan Hou, Zheng Ma, Chunxiao Liu, Chen Change Loy, 2018
- A Generalized Meta-loss function for regression and classification using privileged information, Amina Asif, Muhammad Dawood, Fayyaz ul Amir Afsar Minhas, 2018
- Large scale distributed neural network training through online distillation, Rohan Anil, Gabriel Pereyra, Alexandre Passos, Robert Ormandi, George E. Dahl, Geoffrey E. Hinton, 2018
- KDGAN: Knowledge Distillation with Generative Adversarial Networks, Xiaojie Wang, Rui Zhang, Yu Sun, Jianzhong Qi, 2018
- Deep Face Recognition Model Compression via Knowledge Transfer and Distillation, Jayashree Karlekar, Jiashi Feng, Zi Sian Wong, Sugiri Pranata, 2019
- Relational Knowledge Distillation, Wonpyo Park, Dongju Kim, Yan Lu, Minsu Cho, 2019
- Graph-based Knowledge Distillation by Multi-head Attention Network, Seunghyun Lee, Byung Cheol Song, 2019
- Knowledge Adaptation for Efficient Semantic Segmentation, Tong He, Chunhua Shen, Zhi Tian, Dong Gong, Changming Sun, Youliang Yan, 2019
- Structured Knowledge Distillation for Semantic Segmentation, Yifan Liu, Ke Chen, Chris Liu, Zengchang Qin, Zhenbo Luo, Jingdong Wang, 2019
- Fast Human Pose Estimation, Feng Zhang, Xiatian Zhu, Mao Ye, 2019
- MEAL: Multi-Model Ensemble via Adversarial Learning, Zhiqiang Shen, Zhankui He, Xiangyang Xue, 2019
- Learning Lightweight Lane Detection CNNs by Self Attention Distillation, Yuenan Hou, Zheng Ma, Chunxiao Liu, Chen Change Loy, 2019
- Improved Knowledge Distillation via Teacher Assistant: Bridging the Gap Between Student and Teacher, Seyed-Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Hassan Ghasemzadeh, 2019
- A Comprehensive Overhaul of Feature Distillation, Byeongho Heo, Jeessoo Kim, Sangdoo Yun, Hyojin Park, Nojun Kwak, Jin Young Choi, 2019
- Contrastive Representation Distillation, Yonglong Tian, Dilip Krishnan, Phillip Isola, 2019
- Distillation-Based Training for Multi-Exit Architectures, Mary Phuong, Christoph H. Lampert, Am Campus, 2019
- Learning Metrics from Teachers: Compact Networks for Image Embedding, Lu Yu, Vacit Oguz Yazici, Xialei Liu, Joost van de Weijer, Yongmei Cheng, Arnaud Ramisa, 2019
- On the Efficacy of Knowledge Distillation, Jang Hyun Cho, Bharath Hariharan, 2019
- Revisit Knowledge Distillation: a Teacher-free Framework, Li Yuan, Francis E.H.Tay, Guilin Li, Tao Wang, Jiashi Feng, 2019
- Ensemble Distribution Distillation, Andrey Malinin, Bruno Mlodzeniec, Mark Gales, 2019
- Improving Congruence and Robustness with Noisy Collaboration in Knowledge Distillation, Elahe Arani, Fahad Sarfraz, Bahram Zonooz, 2019
- Self-training with Noisy Student improves ImageNet classification, Qizhe Xie, Eduard Hovy, Minh-Thang Luong, Quoc V. Le, 2019
- Variational Student: Learning Compact and Sparser Networks in Knowledge Distillation Framework, Srinidhi Hegde, Ranjitha Prasad, Ramya Hebbalaguppe, Vishwajith Kumar, 2019
- Preparing Lessons: Improve Knowledge Distillation with Better Supervision, Tiancheng Wen, Shenqi Lai, Xueming Qian, 2019
- Positive-Unlabeled Compression on the Cloud, Yixing Xu, Yunhe Wang, Hanting Chen, Kai Han, Chunjing Xu, Dacheng Tao, Chang Xu, 2019
- Variational Information Distillation for Knowledge Transfer, Sungsoo Ahn, Shell Xu Hu, Andreas Damianou, Neil D. Lawrence, Zhenwen Dai, 2019
- Knowledge Distillation via Instance Relationship Graph, Yufan Liu, Jiajiong Cao, Bing Li, Chunfeng Yuan, Weiming Hu, Yangxi Li and Yunqiang Duan, 2019
- Knowledge Distillation via Route Constrained Optimization, Xiao Jin, Baoyun Peng, Yichao Wu, Yu Liu, Jiaheng Liu, Ding Liang, Junjie Yan, Xiaolin Hu, 2019
- Similarity-Preserving Knowledge Distillation, Frederick Tung, Greg Mori, 2019
- Distilling Object Detectors with Fine-grained Feature Imitation, Tao Wang, Li Yuan, Xiaopeng Zhang, Jiashi Feng, 2019
- Knowledge Squeezed Adversarial Network Compression, Shu Changyong, Li Peng, Xie Yuan, Qu Yanyun, Dai Longquan, Ma Lizhuang, 2019
- Stagewise Knowledge Distillation, Akshay Kulkarni, Navid Panchi, Shital Chidдарwar, 2019
- Knowledge Distillation from Internal Representations, Gustavo Aguilar, Yuan Ling, Yu Zhang, Benjamin Yao, Xing Fan, Edward Guo, 2019
- Knowledge Flow: Improve Upon Your Teachers, Iou-Jen Liu, Jian Peng, Alexander G. Schwing, 2019
- Graph Representation Learning via Multi-task Knowledge Distillation, Jiaqi Ma, Qiaozhu Mei, 2019
- Deep geometric knowledge distillation with graphs, Carlos Lassance, Myriam Bontonou, Ghouthi Boukii Hacene, Vincent Gripon, Jian Tang, Antonio Ortega, 2019
- Correlation Congruence for Knowledge Distillation, Baoyun Peng, Xiao Jin, Jiaheng Liu, Shunfeng Zhou, Yichao Wu, Yu Liu, Dongsheng Li, Zhaoqing Zhang, 2019
- Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation, Linfeng Zhang, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, Kaisheng Ma, 2019
- BAM! Born-Again Multi-Task Networks for Natural Language Understanding, Kevin Clark, Minh-Thang Luong, Urvashi Khandelwal, Christopher D. Manning, Quoc V. Le, 2019
- Self-Knowledge Distillation in Natural Language Processing, Sangchul Hahn, Heeyoul Choi, 2019
- Rethinking Data Augmentation: Self-Supervision and Self-Distillation, Hankook Lee, Sung Ju Hwang, Jinwoo Shin, 2019
- MSD: Multi-Self-Distillation Learning via Multi-classifiers within Deep Neural Networks, Yunteng Luan, Hanyu Zhao, Zhi Yang, Yafei Dai, 2019
- Efficient Video Classification Using Fewer Frames, Shweta Bhardwaj, Mukundhan Srinivasan, Mitesh M. Khapra, 2019
- Retaining Privileged Information for Multi-Task Learning, Fengyi Tang, Cao Xiao, Fei Wang, Jiayu Zhou, Li-Wei Lehman, 2019
- Data-Free Learning of Student Networks, Hanting Chen, Yunhe Wang, Chang Xu, Zhaohui Yang1, Chuanjian Liu, Boxin Shi, Chunjing Xu, Chao Xu, Qi Tian, 2019
- Positive-Unlabeled Compression on the Cloud, Yixing Xu, Yunhe Wang, Hanting Chen, Kai Han, Chunjing Xu, Dacheng Tao, Chang Xu, 2019
- When Does Label Smoothing Help?, Rafael Müller, Simon Kornblith, Geoffrey Hinton, 2019
- TinyBERT: Distilling BERT for Natural Language Understanding, Xiaojie Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, Qun Liu, 2019
- The State of Knowledge Distillation for Classification, Fabian Ruffey, Karanbir Chahal, 2019
- Distilling Task-Specific Knowledge from BERT into Simple Neural Networks, Raphael Tang, Yao Lu, Linqing Liu, Lili Mou, Olga Vechtomova, Jimmy Lin, 2019
- Channel Distillation: Channel-Wise Attention for Knowledge Distillation, Zaida Zhou, Chaoran Zhuge, Xinwei Guan, Wen Liu, 2020
- Residual Knowledge Distillation, Mengya Gao, Yujun Shen, Quanquan Li, Chen Change Loy, 2020
- ResKD: Residual-Guided Knowledge Distillation, Xuewei Li, Songyuan Li, Bourahla Omar, Fei Wu, Xi Li, 2020
- Dreaming to Distill: Data-free Knowledge Transfer via DeepInversion, Hongxu Yin, Pavlo Molchanov, Zhizhong Li, Jose M. Alvarez, Arun Mallya, Derek Hoiem, Niraj K. Jha, Jan Kautz, 2020
- MEAL V2: Boosting Vanilla ResNet-50 to 80%+ Top-1 Accuracy on ImageNet without Tricks, Zhiqiang Shen, Marios Savvides, 2020
- MGD: Matching Guided Distillation, Kaiyu Yue, Jiangfan Deng, Feng Zhou, 2020
- Reducing the Teacher-Student Gap via Spherical Knowledge Distillation, Jia Guo, Minghao Chen, Yao Hu, Chen Zhu, Xiaofei He, Deng Cai, 2020
- Regularizing Class-wise Predictions via Self-knowledge Distillation, Sukmin Yun, Jongjin Park, Kimin Lee, Jinwoo Shin, 2020
- Training data-efficient image transformers & distillation through attention, Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, Hervé Jégou, 2020
- Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks, Lin Wang, Kuk-Jin Yoon, 2020
- Cross-Layer Distillation with Semantic Calibration,Defang Chen, Jian-Ping Mei, Yuan Zhang, Can Wang, Yan Feng, Chun Chen, 2020
- MobileStyleGAN: A Lightweight Convolutional Neural Network for High-Fidelity Image Synthesis, Sergei Belousov, 2021
- Knowledge Distillation: A Survey, Jianping Gou, Baosheng Yu, Stephen John Maybank, Dacheng Tao, 2021
- Refine Myself by Teaching Myself: Feature Refinement via Self-Knowledge Distillation, Mingi Ji, Seungjae Shin, Seunghyun Hwang, Gibeom Park, Il-Chul Moon, 2021
- Complementary Relation Contrastive Distillation, Jinguo Zhu, Shixiang Tang, Dapeng Chen, Shijie Yu, Yakun Liu, Aijun Yang, Mingzhe Rong, Xiaohua Wang, 2021
- Distilling Knowledge via Knowledge Review,Pengguang Chen, Shu Liu, Hengshuang Zhao, Jiaya Jia, 2021
- Hierarchical Self-supervised Augmented Knowledge Distillation, Chuanguang Yang, Zhulin An, Linhang Cai, Yongjun Xu, 2021

Videos

- Dark knowledge, Geoffrey Hinton, 2014
- Model Compression, Rich Caruana, 2016

Implementations

MXNet

- Bayesian Dark Knowledge

PyTorch

- Attention Transfer
- Best of Both Worlds: Transferring Knowledge from Discriminative Learning to a Generative Visual Dialog Model
- Interpreting Deep Classifier by Visual Distillation of Dark Knowledge
- Mean teachers are better role models
- Relational Knowledge Distillation
- Knowledge Transfer via Distillation of Activation Boundaries Formed by Hidden Neurons
- Fast Human Pose Estimation Pytorch
- MEAL: Multi-Model Ensemble via Adversarial Learning
- MEAL-V2: Boosting Vanilla ResNet-50 to 80%+ Top-1 Accuracy on ImageNet without Tricks
- Using Teacher Assistants to Improve Knowledge Distillation
- A Comprehensive Overhaul of Feature Distillation
- Contrastive Representation Distillation
- Transformer model distillation
- TinyBERT
- Data Efficient Model Compression
- Channel Distillation
- Dreaming to Distill: Data-free Knowledge Transfer via DeepInversion
- MGD: Matching Guided Distillation
- torchdistill: A Modular, Configuration-Driven Framework for Knowledge Distillation
- Knowledge Distillation on SSD
- distiller: A large scale study of Knowledge Distillation
- Knowledge-Distillation-Zoo: Pytorch implementation of various Knowledge distillation (KD) methods
- A PyTorch implementation for exploring deep and shallow knowledge distillation (KD) experiments with flexibility
- Neural Network Distiller by Intel AI Lab: a Python package for neural network compression research.
- KD_Lib : A Pytorch Knowledge Distillation library for benchmarking and extending works in the domains of Knowledge Distillation, Pruning, and Quantization.
- Vision Transformer Distillation
- Cross-Layer Distillation with Semantic Calibration
- Refine Myself by Teaching Myself: Feature Refinement via Self-Knowledge Distillation
- Distilling Knowledge via Knowledge Review
- Hierarchical Self-supervised Augmented Knowledge Distillation

Lua

- Example for teacher/student-based learning

Torch

- Distilling knowledge to specialist ConvNets for clustered classification
- Sequence-Level Knowledge Distillation, Neural Machine Translation on Android
- cifar.torch distillation
- ENet-SAD

Theano

- FitNets: Hints for Thin Deep Nets
- Transfer knowledge from a large DNN or an ensemble of DNNs into a small DNN

Lasagne + Theano

- Experiments-with-Distilling-Knowledge

Tensorflow

- Deep Model Compression: Distilling Knowledge from Noisy Teachers
- Distillation
- An example application of neural network distillation to MNIST
- Data-free Knowledge Distillation for Deep Neural Networks
- Inspired by net2net, network distillation
- Deep Reinforcement Learning, knowledge transfer
- Knowledge Distillation using Tensorflow
- Knowledge Distillation Methods with Tensorflow
- Zero-Shot Knowledge Distillation in Deep Networks in ICML2019
- Knowledge_distillation_benchmark via Tensorflow2.0

Caffe

- Face Model Compression by Distilling Knowledge from Neurons
- KnowledgeDistillation Layer (Caffe implementation)
- Knowledge distillation, realized in caffe
- Cross Modal Distillation for Supervision Transfer
- Multi-Label Image Classification via Knowledge Distillation from Weakly-Supervised Detection
- Knowledge Distillation via Instance Relationship Graph

Keras

- Knowledge distillation with Keras
- keras google-vision's distillation
- Distilling the knowledge in a Neural Network

- https://github.com/PaddlePaddle/PaddleSlim/tree/release/2.0.0/docs/zh_cn/tutorials

Awesome Knowledge Distillation

deep-learning

knowledge-distillation

teacher-student

knowledge-transfer

co-training

model-compression

distillation

kd

knowledge-distillation

distillation-model

model-distillation

Readme

Apache-2.0 License

2.7k stars

111 watching

449 forks

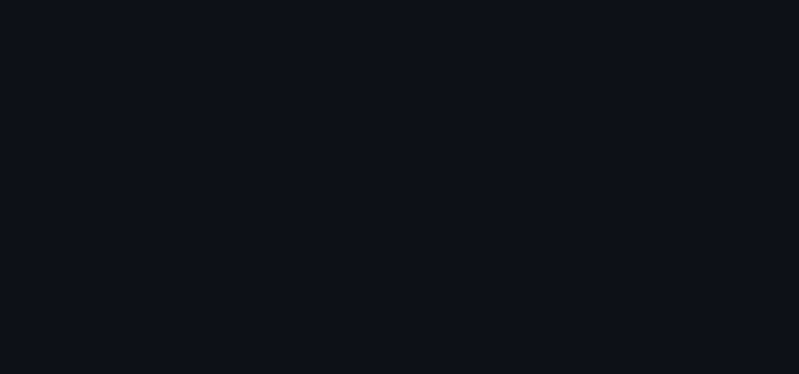
Releases

No releases published

Packages

No packages published

Contributors 21



+ 10 contributors