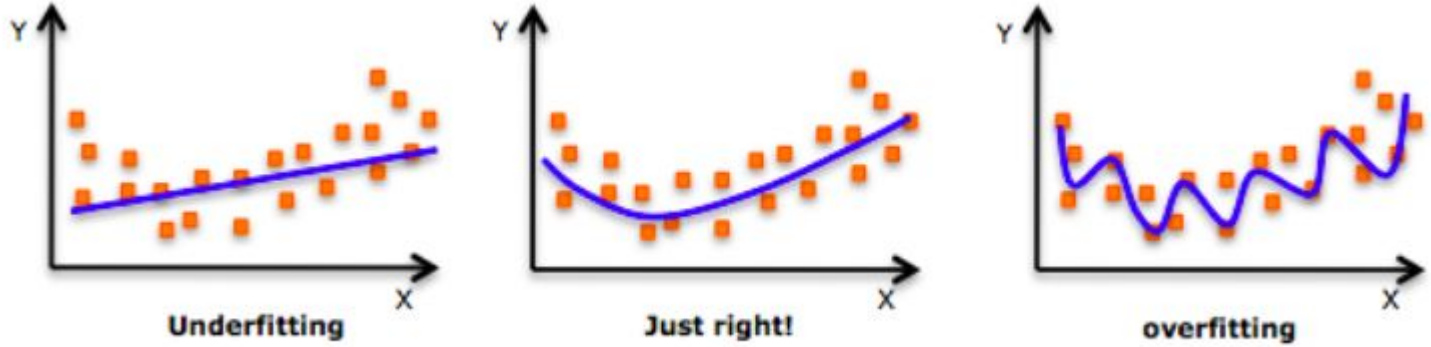
Post-Regularization

Yara Mohamadi

Supervisor: Dr. Sadeghi

Regularization



Regularization techniques

Explicit methods:

- L1 & L2 regularization

Implicit methods:

- Ensembling
- Data augmentation and synthesis
- Early stopping
- Dropout
- Batchnorm
- Constraints on the model (E.g, CNNs)
- ...

Why post-regularization

- To fix an over-fit model without retraining from scratch
- To speed up hyperparameter finding
- To easily change regularization intensity according to our needs
- To take another step towards general AI !

Types of post-regularization

- Modified training (Model dependent)
  - Controlling the regularization parameter in inference
- Regular training
  - Starting the regularization process after some iterations (Model dependent)
  - Model agnostic post-training regularization
  - Post-training manipulation of the model (Model dependent)

Some have done it...

A hand-full of Post-Regularization inspired works exist in various fields:

- Imposing orthogonality
- Using ensemble models
- Continual Learning
- Curriculum learning
- Self-distillation
- Robustness techniques
- Dropout
- Pruning
- Quantization
- Optimization methods
- Test-time augmentations
- ...

However, most of them do not have my perspective...

Obtaining Adjustable Regularization for Free via Iterate Averaging

Search... Expand

Origin paper

Obtaining Adjustable Regularization for Free via Iterate Averaging

Jingfeng Wu, Vladimir Braverman, Lin F. Yang 2020

Which Algorithmic Choices Matter at Which Batch Sizes? Insights From a Noisy Quadratic...

Guodong Zhang, Lala Li, Zachary Nado, James... 2019

An Empirical Study of Large-Batch Stochastic Gradient Descent with Structured Covariance...

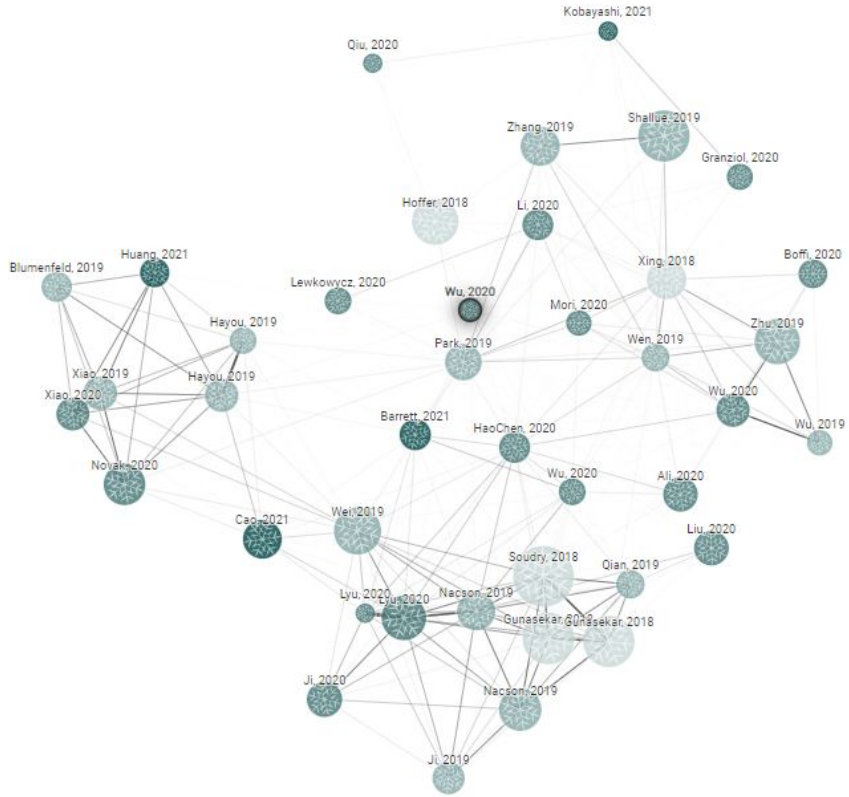
Yeming Wen, Kevin Luk, M. Gazeau, Guodong... 2019

Mean-field Behaviour of Neural Tangent Kernel for Deep Neural Networks

Soufiane Hayou, A. Doucet, J. Rousseau 2019

Lexicographic and Depth-Sensitive Margins in Homogeneous and Non-Homogeneous Deep...

Mor Shpigel Nacson, Suriya Gunasekar, J. Lee,... 2019



Snow On

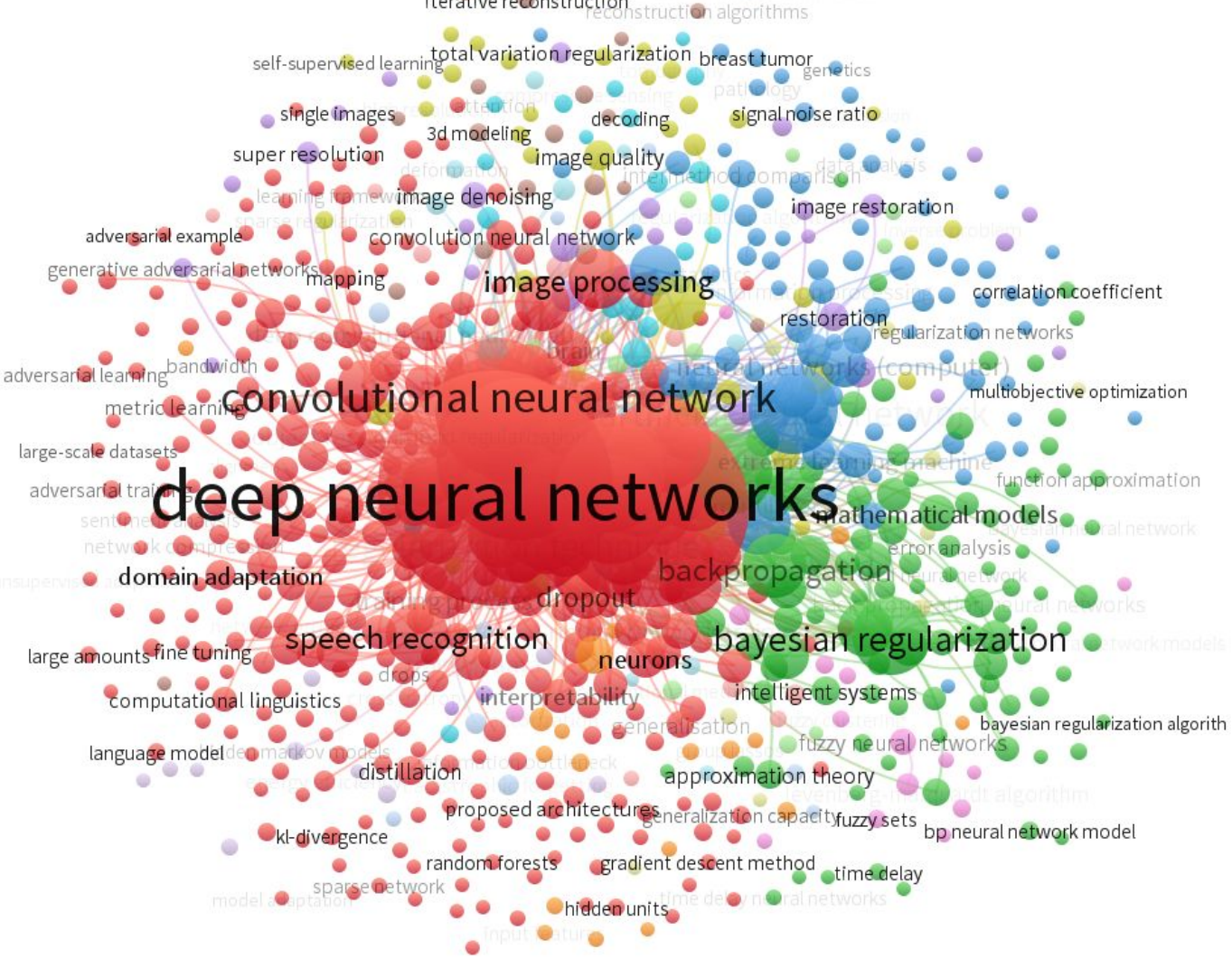
Created on Aug 21 2021 Building new...

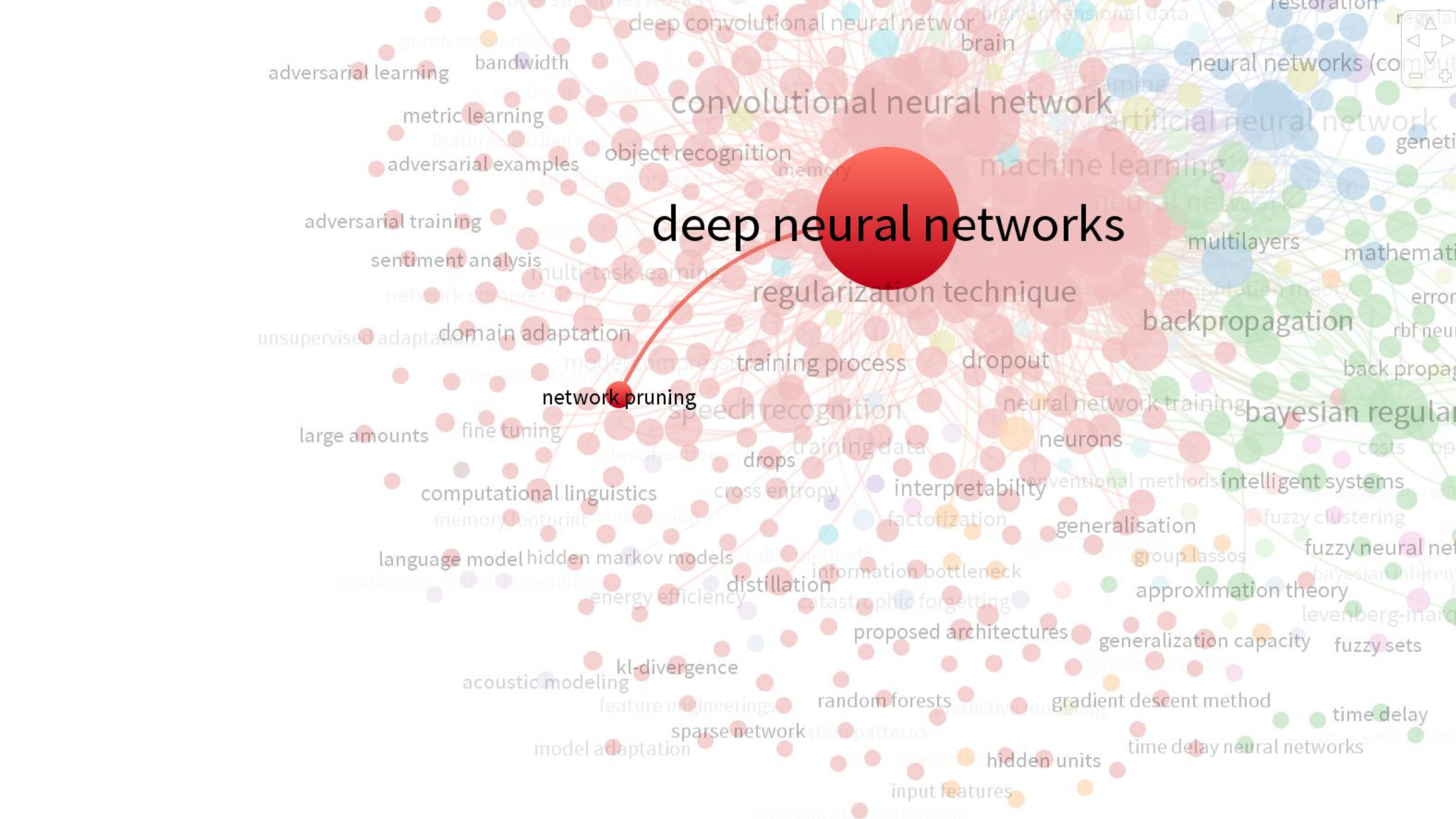
2018 2021

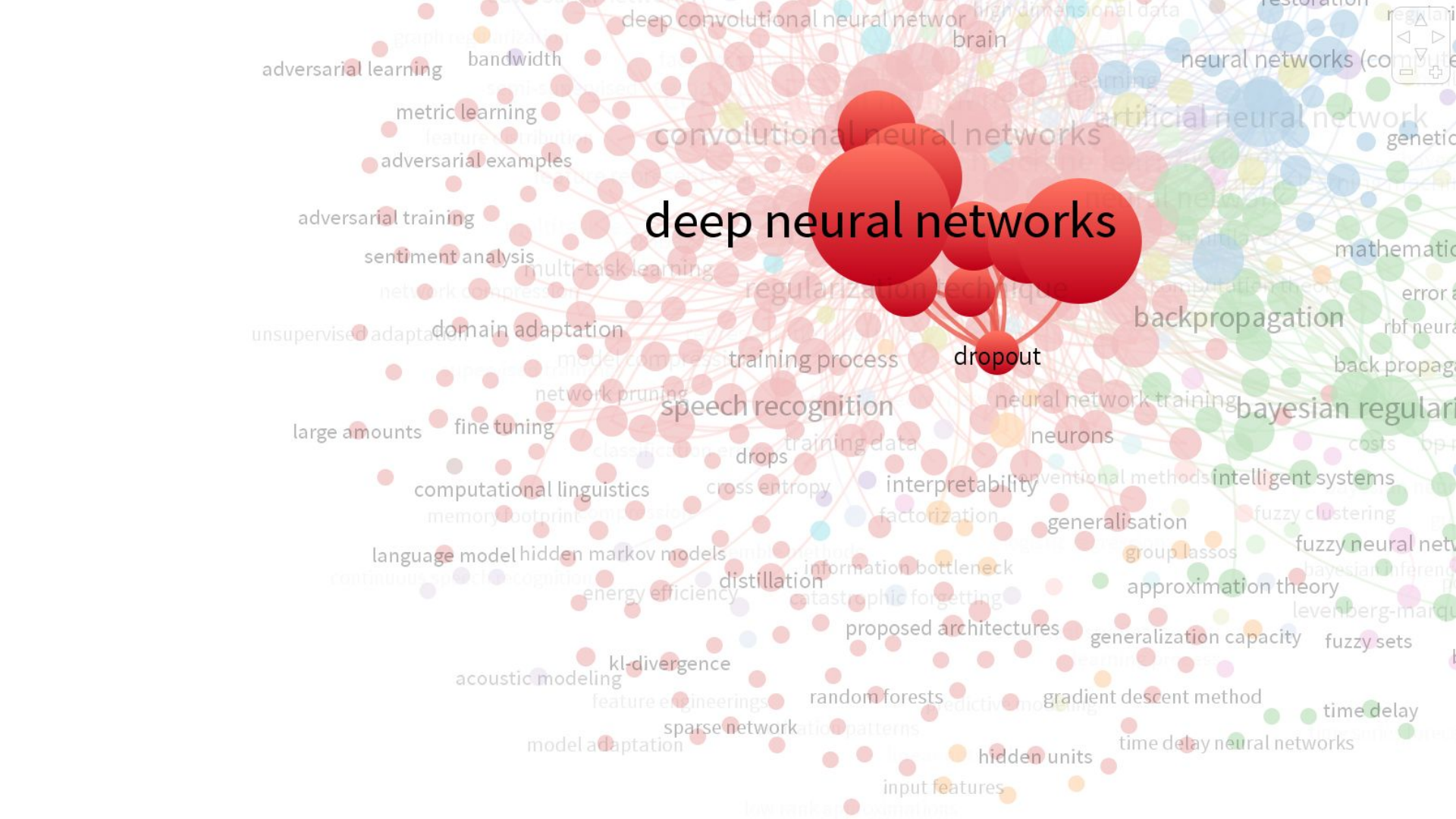
First Search (scopus)

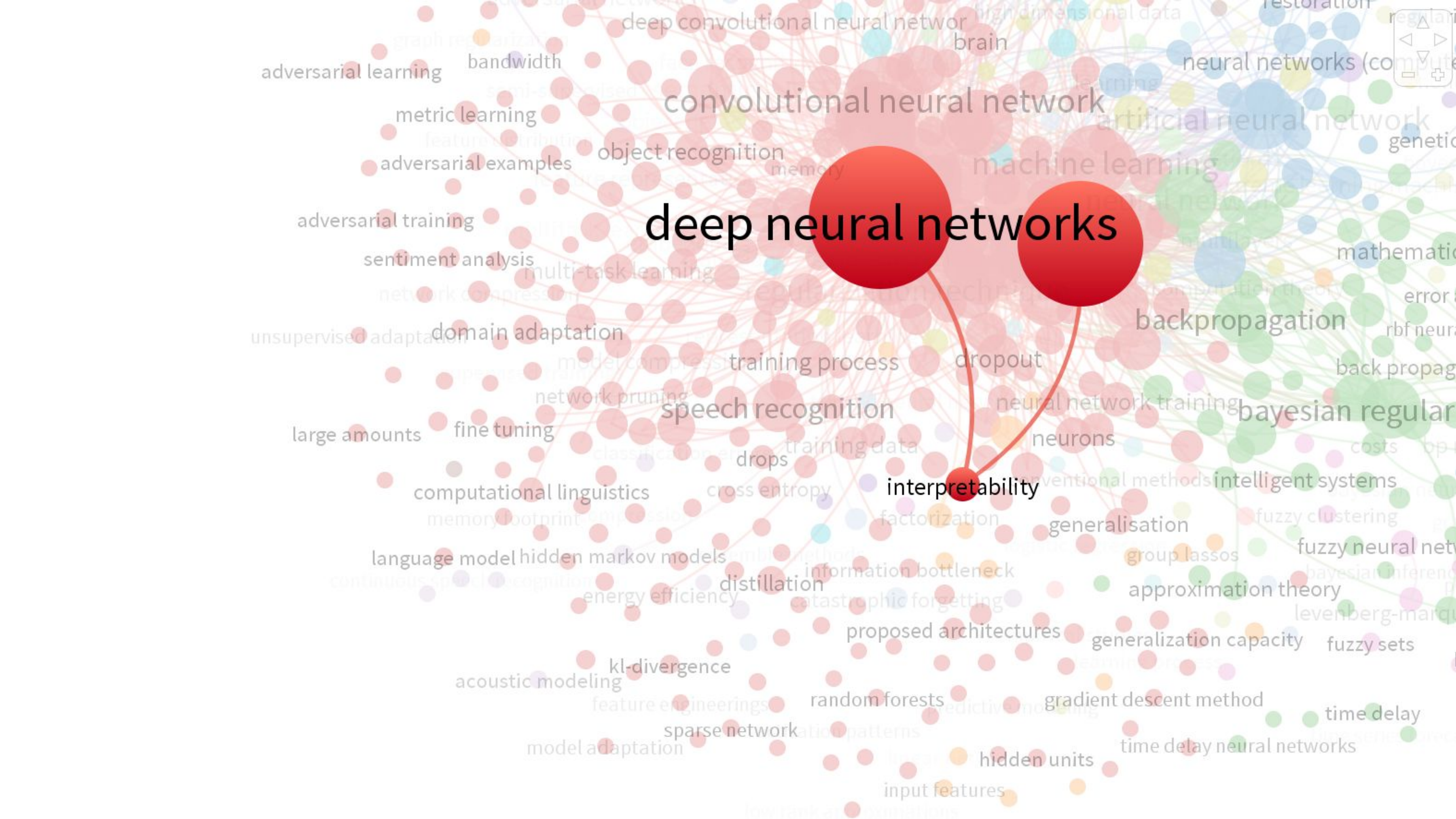
8,543 document results

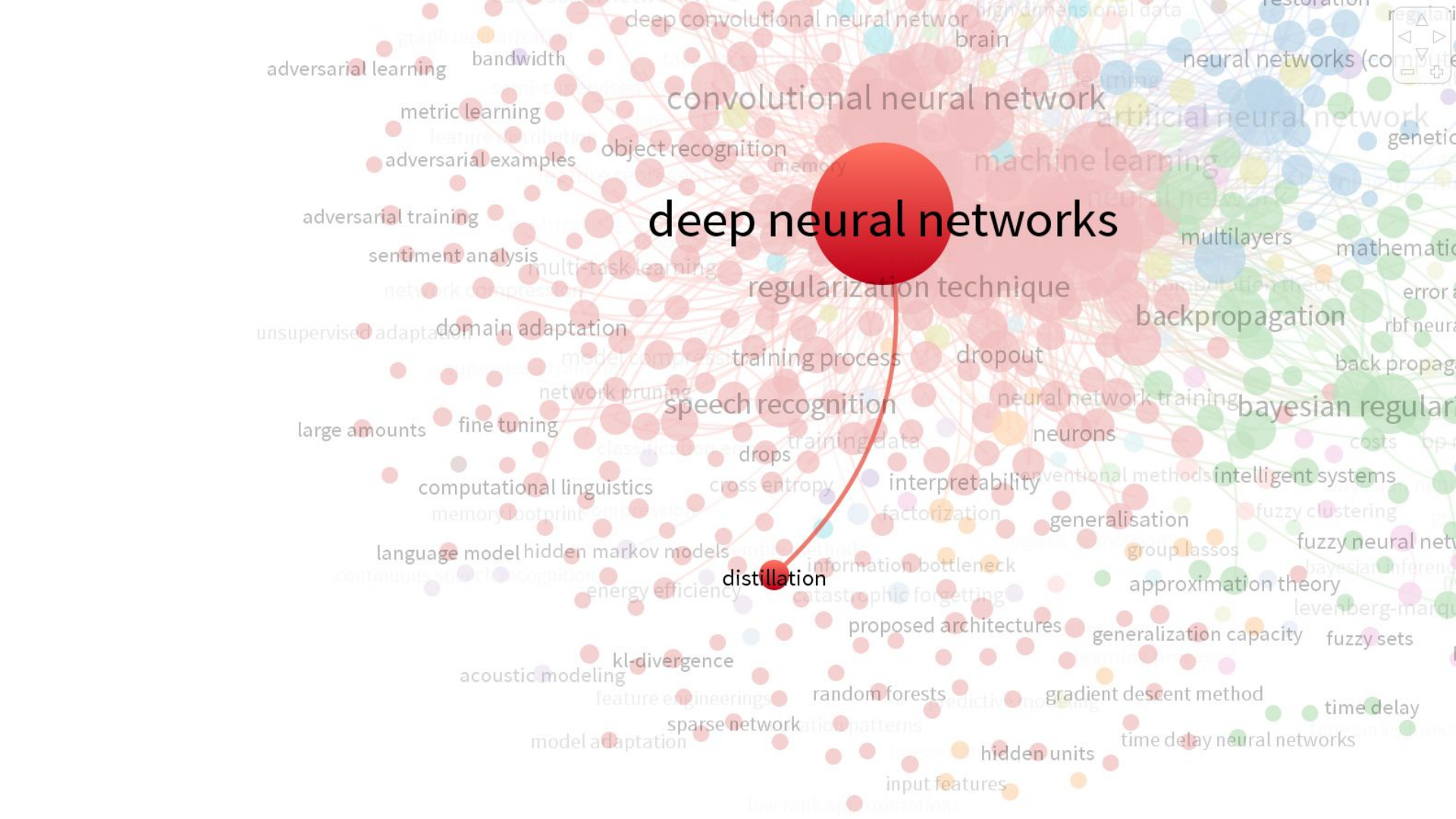
(TITLE-ABS-KEY ({Regularization}) AND TITLE-ABS-KEY ({machine learning} OR "neural networks" OR "deep neural networks"))

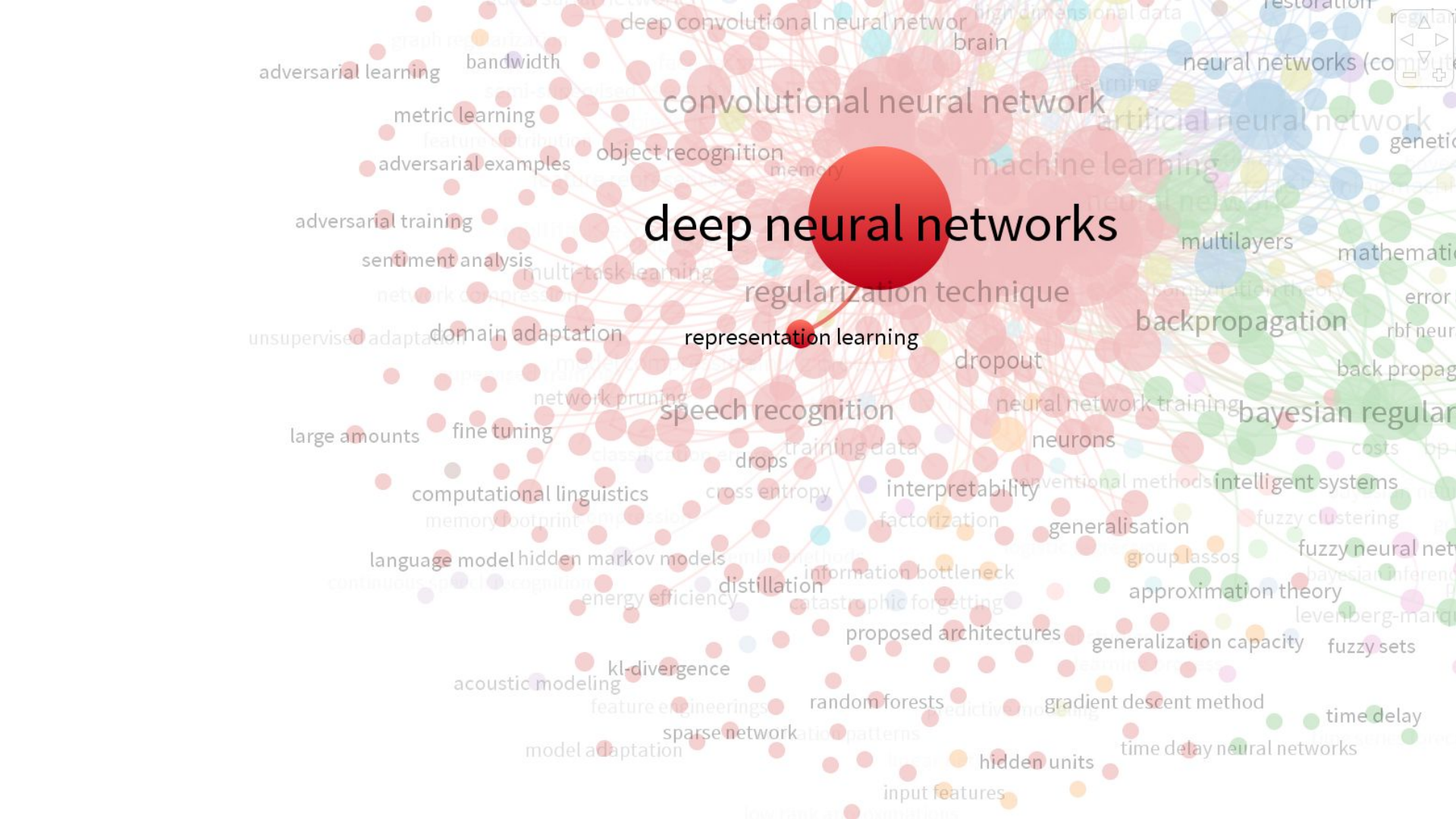

deep neural networks
convolutional neural network
image processing
backpropagation
bayesian regularization
speech recognition
dropout
domain adaptation
mathematical models
neurons
image denoising
image restoration
convolution neural network
total variation regularization
breast tumor
signal noise ratio
decoding
3d modeling
image quality
super resolution
single images
self-supervised learning
adversarial example
generative adversarial networks
mapping
correlation coefficient
restoration
regularization networks
multiobjective optimization
adversarial learning
bandwidth
metric learning
large-scale datasets
function approximation
error analysis
large amounts
fine tuning
computational linguistics
interpretability
intelligent systems
bayesian regularization algorith
fuzzy neural networks
language model
distillation
approximation theory
proposed architectures
generalization capacity
fuzzy sets
bp neural network model
kl-divergence
random forests
gradient descent method
time delay
sparse network
hidden units

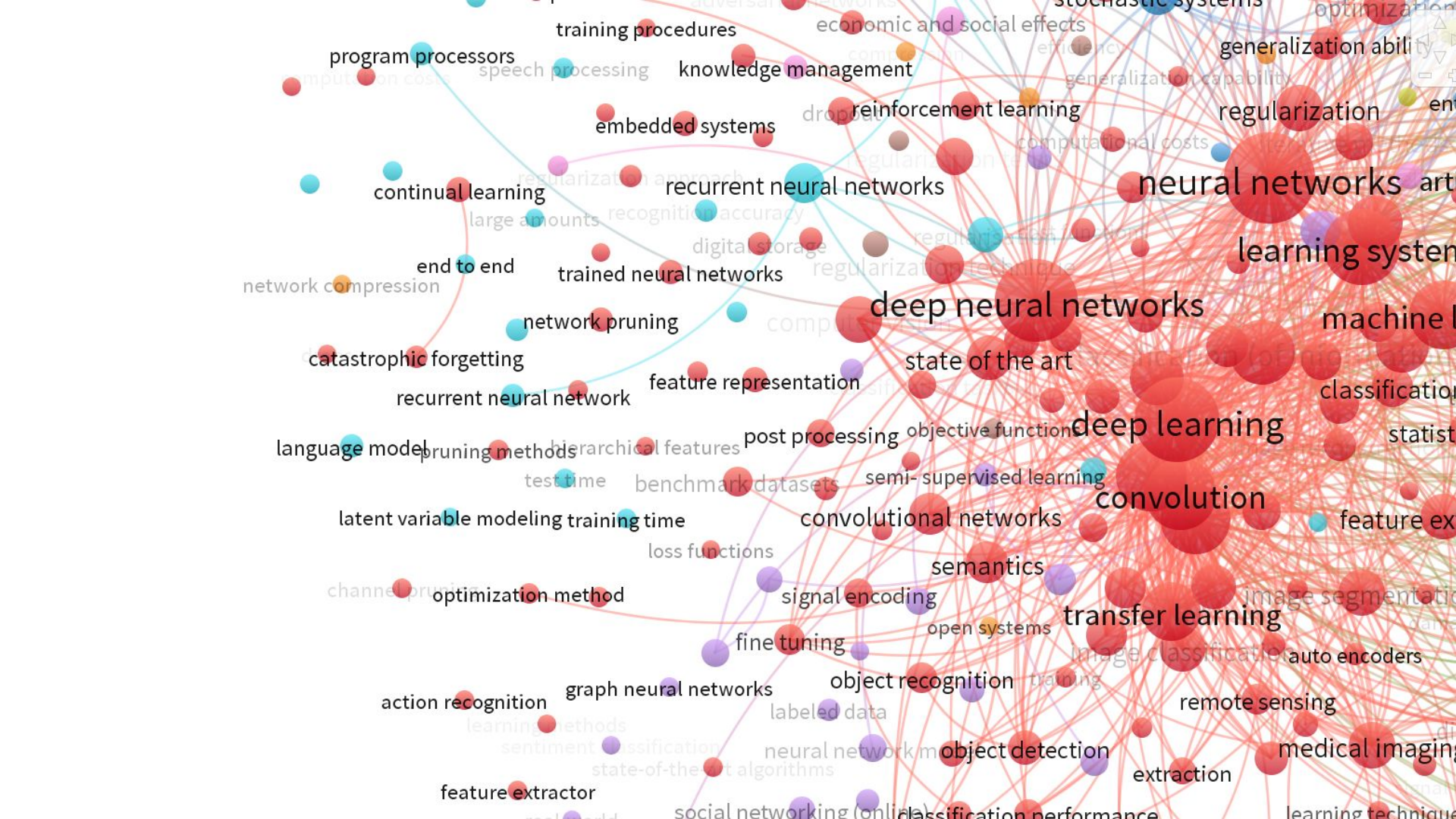

deep neural networks
network pruning
convolutional neural network
deep convolutional neural networ
brain
neural networks (co
artificial neural network
machine learning
object recognition
regularization technique
multilayers
backpropagation
training process
dropout
speech recognition
neural network training
bayesian regular
neurons
adversarial learning
bandwidth
metric learning
adversarial examples
adversarial training
sentiment analysis
multi-task learning
domain adaptation
large amounts
fine tuning
drops
interpretability
computational linguistics
cross entropy
factorization
generalisation
intelligent systems
fuzzy clustering
fuzzy neural net
language model
hidden markov models
information bottleneck
group lassos
distillation
energy efficiency
approximation theory
proposed architectures
generalization capacity
fuzzy sets
kl-divergence
acoustic modeling
random forests
gradient descent method
time delay
sparse network
model adaptation
time delay neural networks
hidden units
input features
mathemati
error
rbf neu
back propag
geneti
restoration


deep convolutional neural networ
brain
adversarial learning
bandwidth
neural networks (co
metric learning
convolutional neural networks
artificial neural network
adversarial examples
adversarial training
deep neural networks
sentiment analysis
multi-task learning
regularization technique
backpropagation
domain adaptation
unsupervised adaptation
training process
dropout
network pruning
speech recognition
neural network training
large amounts
fine tuning
neurons
training data
drops
computational linguistics
cross entropy
interpretability
intelligent systems
factorization
generalisation
language model
hidden markov models
group lassos
fuzzy neural net
information bottleneck
distillation
approximation theory
energy efficiency
proposed architectures
generalization capacity
fuzzy sets
kl-divergence
acoustic modeling
random forests
gradient descent method
time delay
sparse network
time delay neural networks
model adaptation
hidden units
input features


deep neural networks
interpretability
convolutional neural network
machine learning
artificial neural network
brain
adversarial learning
bandwidth
metric learning
adversarial examples
object recognition
adversarial training
sentiment analysis
multi-task learning
domain adaptation
training process
dropout
backpropagation
speech recognition
network pruning
neural network training
large amounts
fine tuning
neurons
drops
training data
computational linguistics
cross entropy
intelligent systems
factorization
generalisation
language model
hidden markov models
group lassos
information bottleneck
distillation
energy efficiency
approximation theory
catastrophic forgetting
proposed architectures
generalization capacity
fuzzy sets
kl-divergence
acoustic modeling
random forests
gradient descent method
time delay
sparse network
model adaptation
time delay neural networks
hidden units
input features
fuzzy clustering


deep convolutional neural networ
brain
adversarial learning
bandwidth
neural networks (co
convolutional neural network
metric learning
artificial neural network
genetic
object recognition
adversarial examples
machine learning
adversarial training
deep neural networks
multilayers
sentiment analysis
multi-task learning
regularization technique
error
backpropagation
domain adaptation
rbf neura
training process
dropout
back propag
network pruning
speech recognition
neural network training
bayesian regular
large amounts
fine tuning
neurons
drops
intelligent systems
computational linguistics
cross entropy
interpretability
fuzzy clustering
generalisation
fuzzy neural net
language model
hidden markov models
group lassos
information bottleneck
distillation
approximation theory
energy efficiency
proposed architectures
generalization capacity
fuzzy sets
kl-divergence
acoustic modeling
random forests
gradient descent method
time delay
sparse network
time delay neural networks
model adaptation
hidden units
input features


deep neural networks
representation learning
regularization technique
convolutional neural network
deep convolutional neural networ
brain
neural networks (co
adversarial learning
bandwidth
metric learning
adversarial examples
object recognition
machine learning
adversarial training
sentiment analysis
multi-task learning
multilayers
mathematic
error
backpropagation
rbf neur
domain adaptation
dropout
back propag
speech recognition
neural network training
bayesian regular
large amounts
fine tuning
neurons
drops
intelligent systems
computational linguistics
interpretability
factorization
generalisation
fuzzy clustering
fuzzy neural net
language model
hidden markov models
group lassos
information bottleneck
distillation
approximation theory
energy efficiency
proposed architectures
generalization capacity
fuzzy sets
kl-divergence
acoustic modeling
random forests
gradient descent method
time delay
sparse network
time delay neural networks
hidden units
input features


training procedures
program processors
speech processing
knowledge management
economic and social effects
generalization ability
reinforcement learning
regularization
embedded systems
computational costs
continual learning
recurrent neural networks
neural networks
large amounts
digital storage
learning system
end to end
trained neural networks
network compression
deep neural networks
network pruning
machine
catastrophic forgetting
state of the art
feature representation
recurrent neural network
classification
deep learning
post processing
objective function
language model
pruning methods
hierarchical features
test time
benchmark datasets
semi- supervised learning
convolution
latent variable modeling
training time
convolutional networks
feature ex
loss functions
semantics
optimization method
signal encoding
image segmentation
transfer learning
open systems
fine tuning
auto encoders
object recognition
graph neural networks
action recognition
remote sensing
labeled data
object detection
medical imaging
extraction
feature extractor
social networking (online)
classification performance

Second search

2,539 document results

(TITLE-ABS-KEY (
"regularization" OR "post-regularization" OR "generalization" OR "robustness" OR "robust" OR "adversarial" OR "iterate averaging" OR "slimming" OR "augmentation" OR "pruning" OR "quantization" OR "compression" OR "dropout" OR "distillation" OR "orthogonal" OR {double descent} OR "grokking" OR "curriculum" OR {continual learning} OR {catastrophic forgetting} OR "overfit" OR "over-fit")

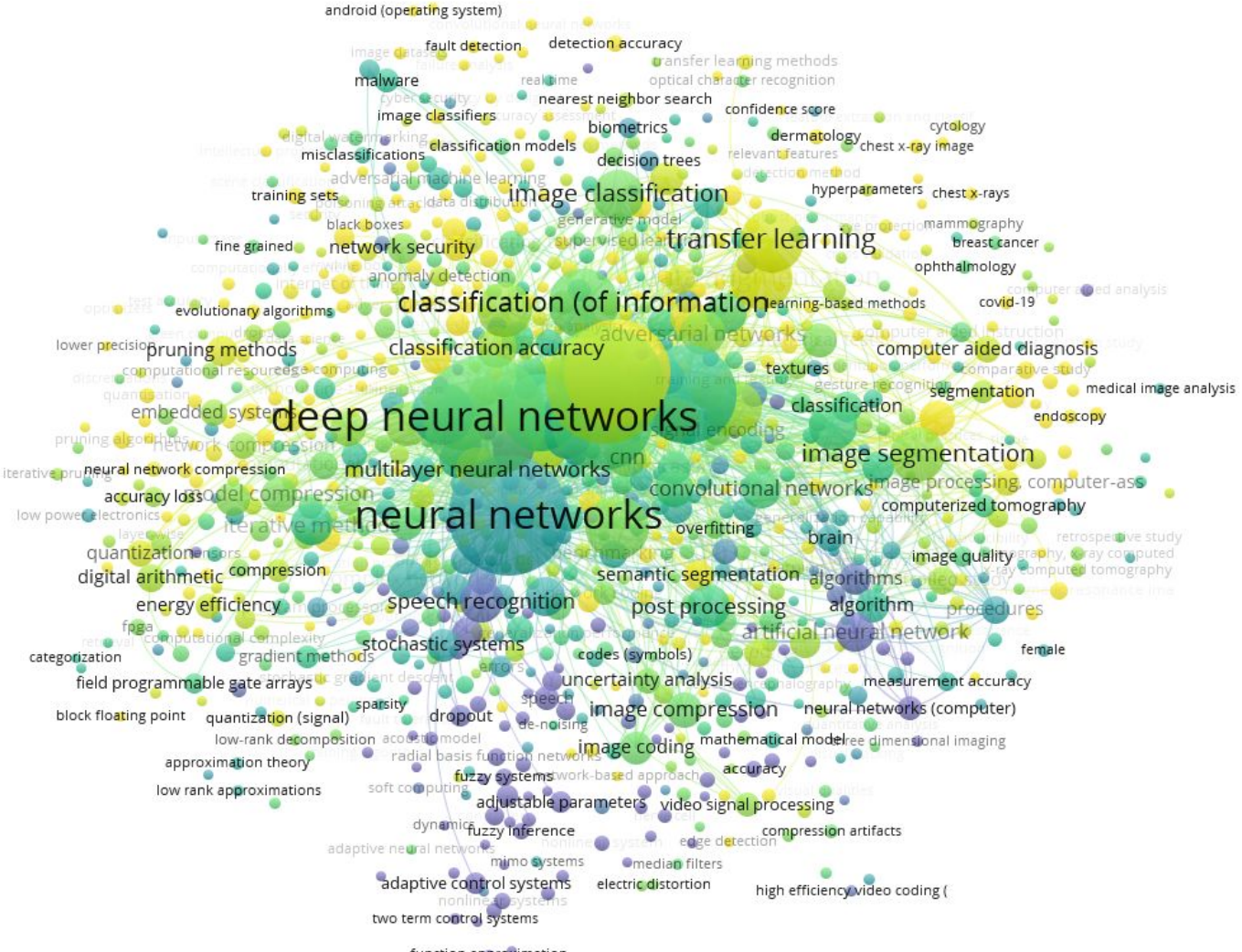
AND TITLE-ABS-KEY (
"machine learning" OR "neural networks" OR "deep neural networks")

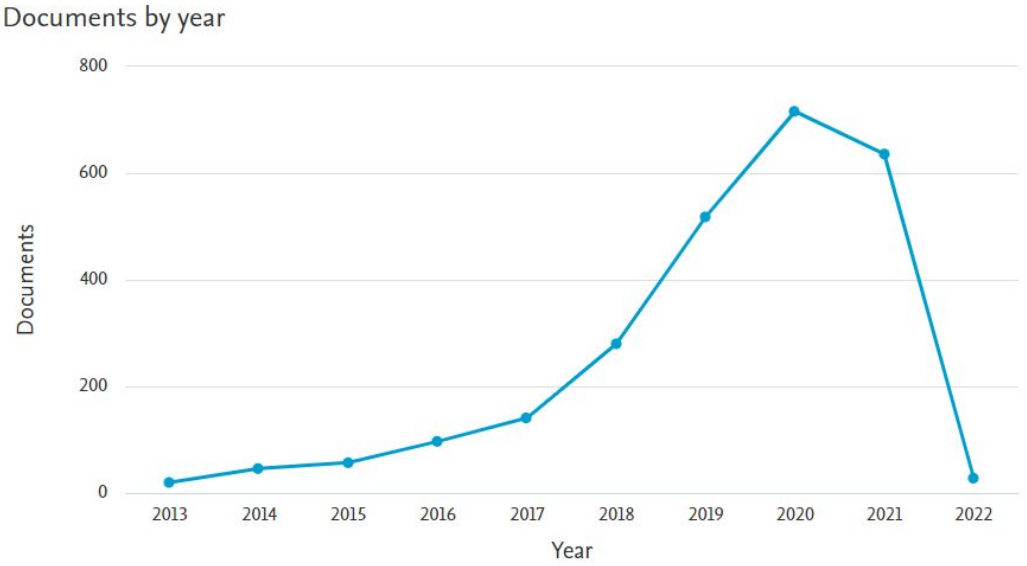
AND TITLE-ABS-KEY (
"post-training" OR "test-time" OR "post-learning" OR "adjustable" OR "disentangled" OR "post-processing" OR "retraining" OR "fine-tuning" OR {after training} OR {after learning}))

Second search

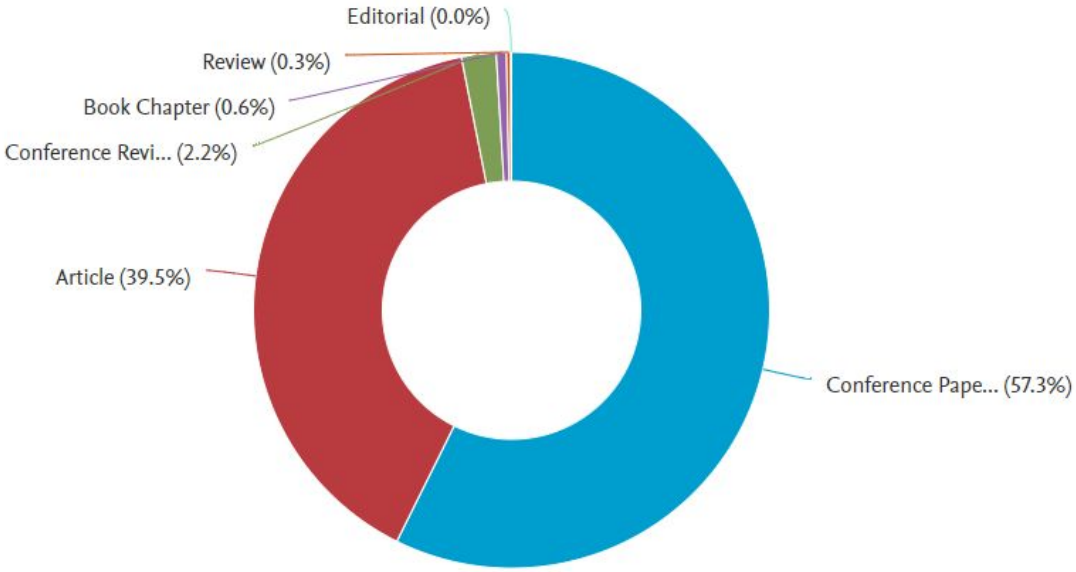
2,539 document results

AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013)) AND (LIMIT-TO (LANGUAGE , "English"))

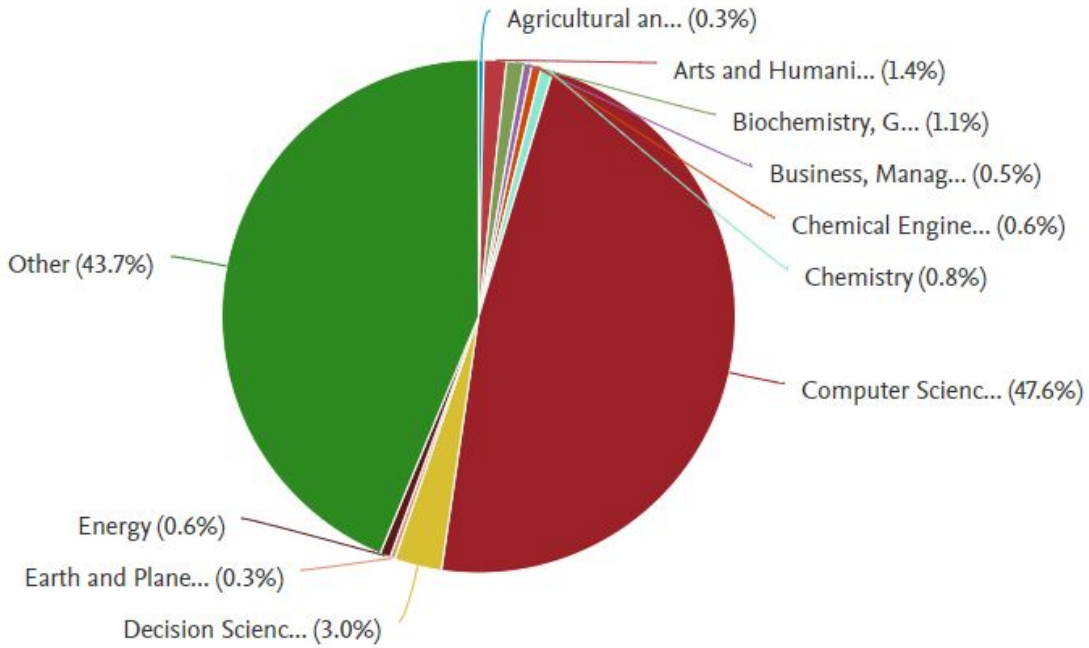

android (operating system)
fault detection
detection accuracy
malware
nearest neighbor search
confidence score
image classifiers
biometrics
dermatology
cytology
misclassifications
classification models
chest x-ray image
decision trees
image classification
hyperparameters
chest x-rays
training sets
black boxes
transfer learning
mammography
fine grained
network security
breast cancer
anomaly detection
ophthalmology
classification (of information
covid-19
evolutionary algorithms
adversarial networks
lower precision
pruning methods
classification accuracy
computer aided diagnosis
textures
medical image analysis
segmentation
embedded systems
deep neural networks
classification
endoscopy
image segmentation
neural network compression
multilayer neural networks
cnn
convolutional networks
accuracy loss
model compression
computerized tomography
neural networks
overfitting
brain
quantization
image quality
digital arithmetic
compression
semantic segmentation
algorithms
energy efficiency
speech recognition
post processing
algorithm
procedures
artificial neural network
stochastic systems
female
categorization
gradient methods
codes (symbols)
field programmable gate arrays
uncertainty analysis
measurement accuracy
image compression
neural networks (computer)
block floating point
quantization (signal)
sparsity
dropout
speech
de-noising
low-rank decomposition
mathematical model
image coding
approximation theory
accuracy
fuzzy systems
low rank approximations
adjustable parameters
video signal processing
dynamics
fuzzy inference
compression artifacts
mimo systems
median filters
adaptive control systems
electric distortion
high efficiency video coding (
two term control systems
function approximation


Documents by year
800
600
400
200
0
Documents
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year

Documents by type

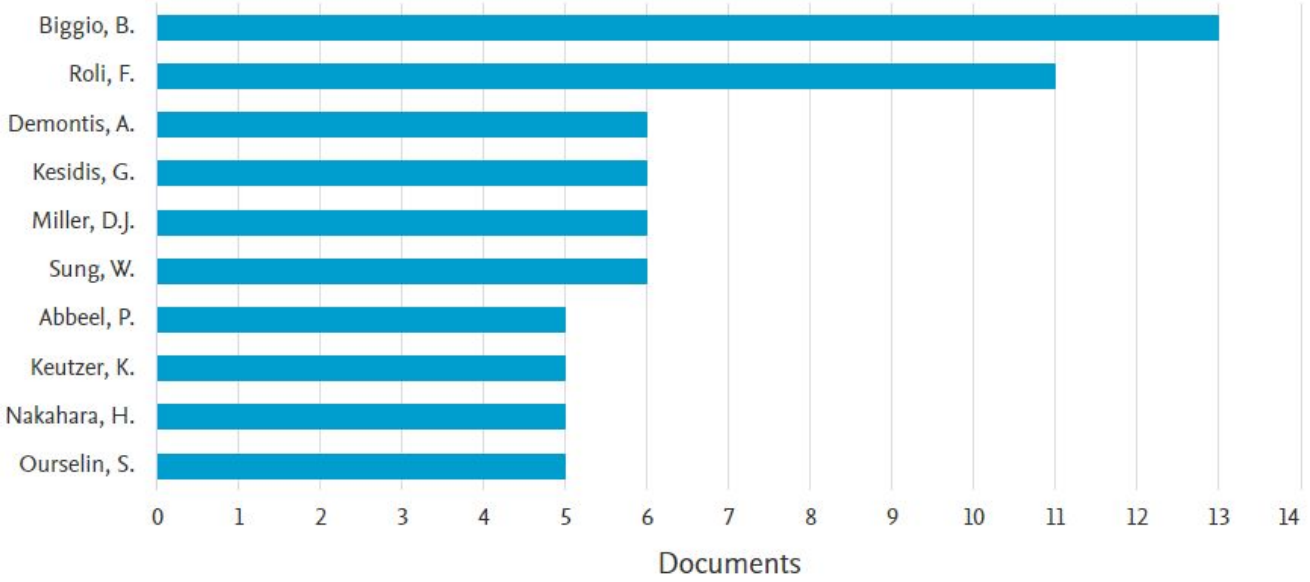


Documents by subject area



Documents by author

Compare the document counts for up to 15 authors.



Thank you.