Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Geoffrey Hinton

Emeritus Prof. Computer Science, University of Toronto

[Google Scholar profile]

Generated 2026-05-25 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

8,539	10,109	710	190
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

95.0% independent of 5,250 classified citing papers

Citation type	Count
Independent	4,990
Self-citation	18
Co-author	196
Same-institution	46

3,289 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher introduced Dropout as a foundational regularization technique for neural networks, subsequently advancing the field through seminal work on knowledge distillation and comprehensive deep learning synthesis.

The researcher's contribution centers on the development and dissemination of core techniques in deep learning, anchored by the 2014 JMLR paper 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting.' This work established a critical method for improving model generalization, serving as the foundation for subsequent influential publications.

This line of work appears to address the persistent challenge of overfitting in complex neural architectures. By introducing Dropout, the researcher provided a simple yet effective solution that enabled more robust training. The follow-up papers, including 'Distilling the Knowledge in a Neural Network' (2015) and the comprehensive review 'Deep learning' (2015), suggest a broader effort to optimize network efficiency and consolidate theoretical understanding within the community.

The significance of this research is evidenced by its extensive adoption, with the core paper accumulating 62,372 citations and the follow-up works garnering 31,768 and 111,795 citations respectively. Furthermore, citation analysis reveals that 98.4% of citing papers originate from independent researchers, indicating that these contributions have become standard, widely utilized tools across the global machine learning community rather than niche academic exercises.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 1,860 · **91 flagged influential by Semantic Scholar**

CORE PAPER

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

2014 · Journal of Machine Learning Research (JMLR) · 62,372 citations (GS)

Field-normalised: 42,929 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2014 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	iCaRL: Incremental Classifier and Representation Learning (2017)	Institute of Science and Technology Austria, IST Austria	Austria	Methodology
2	Analyzing and Improving the Training Dynamics of Diffusion Models	NVIDIA	United States	—
3	Segmenter: Transformer for Semantic Segmentation	Inria, École normale supérieure, CNRS, PSL Research University	France	—
4	Self-regulating prompts: Foundational model adaptation without forgetting (2023)	Mohamed bin Zayed University of AI, Mohamed bin Zayed University of Artificial Intelligence, University of California, Merced	United Arab Emirates, United States	Methodology
5	Segment Anything	Meta	—	—
6	SimCSE: Simple Contrastive Learning of Sentence Embeddings (2021)	Princeton University, Tsinghua University	China, United States	Methodology
7	mixup: Beyond Empirical Risk Minimization (2017)	Facebook, Facebook AI Research, Massachusetts Institute of Technology	United States	Methodology
8	How Powerful are Graph Neural Networks? (2018)	MIT, Stanford University	United States	Background

No.	Citing paper	Citing institution(s)	Country	S2
9	Activation functions: Comparison of trends in practice and research for deep learning (2018)	University of Strathclyde	United Kingdom	Methodology
10	On the Convergence of Adam and Beyond (2019)	Carnegie Mellon University, Google Research	United States	Influential
11	Hyper-Parameter Optimization: A Review of Algorithms and Applications (2020)	Inspur, Inspur Electronic Information Industry Co., Ltd	China	Methodology
12	YOLOv4: Optimal Speed and Accuracy of Object Detection (2020)	Academia Sinica, Independent Researcher	Taiwan	—
13	No Language Left Behind: Scaling Human-Centered Machine Translation (2024)	Meta, Meta AI	—	—
14	eDiff-I: Text-to-Image Diffusion Models with an Ensemble of Expert Denoisers (2023)	NVIDIA, NVIDIA Corporation	United States	Methodology
15	A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU (2023)	Universiti Putra Malaysia, University Putra Malaysia	Malaysia	—
16	The Falcon Series of Open Language Models	Technology Innovation Institute	United Arab Emirates	—
17	Mechanistic Interpretability for AI Safety – A Review (2024)	University of Amsterdam	Netherlands	Background
18	Chameleon: Mixed-modal early-fusion foundation models (2024)	Meta	—	Methodology
19	From System 1 to System 2: A Survey of Reasoning Large Language Models (2025)	AiShiWeiLai AI Research, Chinese Academy of Sciences, City University of Hong Kong and the Hong Kong University of Science and Technology (Guangzhou)	China, United Arab Emirates, United Kingdom	—
20	Graph Neural Networks for Social Recommendation (2019)	City University of Hong Kong, Michigan State University, The Hong Kong Polytechnic University	China, Hong Kong, United States	Methodology
21	A Survey of Deep Active Learning (2021)	Carnegie Mellon University, Monash University, National Institute of Technology Kurukshetra	Australia, China, India	Methodology
22	Trustworthy AI: From Principles to Practices (2022)	—	—	—
23	Heterogeneous Federated Learning: State-of-the-art and Research Challenges	Hong Kong Baptist University, Nanyang Technological University, Wuhan University	China, Singapore	Methodology
24	A Comprehensive Overview of Large Language Models (2025)	Australian National University, King Fahd University of Petroleum and Minerals, The Chinese University of Hong Kong	Australia, China, Pakistan	—

No.	Citing paper	Citing institution(s)	Country	S2
25	VM-UNet: Vision Mamba UNet for Medical Image Segmentation (2025)	Shanghai Jiao Tong University	China	—
26	A Comprehensive Survey of Continual Learning: Theory, Method and Application	Tsinghua University	China	Background
27	A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications (2024)	University of Huddersfield	United Kingdom	Background
28	Deep Learning for Health Informatics (2016)	Imperial College London	United Kingdom	**Methodology**
29	Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review (2017)	University of South Africa	South Africa	**Methodology**
30	Multimodal Machine Learning: A Survey and Taxonomy (2018)	Carnegie Mellon University	United States	—

Showing the 30 most-cited of 803 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY iCaRL: Incremental Classifier and Representation Learning

"As a consequence, standard end-to-end learning methods can be used, such as backpropagation with mini-batches, but also recent improvements, such as dropout [38], adaptive stepsize selection [14] or batch nor-"

METHODOLOGY Self-regulating prompts: Foundational model adaptation without forgetting

"This category includes methods such as data augmentations [52, 55, 5], dropout [42], model ensembling [18, 47], label smoothing [43] and batch normalization [19]."

METHODOLOGY SimCSE: Simple Contrastive Learning of Sentence Embeddings

"Our unsupervised SimCSE simply predicts the input sentence itself with only dropout (Srivastava et al., 2014) used as noise (Figure 1(a))."

METHODOLOGY mixup: Beyond Empirical Risk Minimization

"Dropout (Srivastava et al., 2014) is considered the state-of-the-art method for learning with corrupted labels (Arpit et al., 2017)."

METHODOLOGY Activation functions: Comparison of trends in practice and research for deep learning

"for model improvement for DL algorithms exist in literature which includes the use of batch-normalisation and regularisation, [9], [10], [11], dropout [9], proper initialisation [12], good choice of AF to mention a few [10], [12], [13], [14]."

FOLLOW-UP WORK

Distilling the Knowledge in a Neural Network

2015 · NIPS 2014 Deep Learning and Representation Learning Workshop · 31,768 citations (GS)

Field-normalised: 23,776 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks (2023)	Hong Kong University of Science and Technology, Rutgers University, Texas State University	Hong Kong, United States	**Methodology**
2	SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy (2023)	East China Normal University, Tongji University	China	—

No.	Citing paper	Citing institution(s)	Country	S2
3	EfficientSAM: Leveraged Masked Image Pre-training for Efficient Segment Anything (2024)	Meta	—	Background
4	One-step Diffusion with Distribution Matching Distillation	Adobe, Adobe Research, Massachusetts Institute of Technology	United States	Methodology
5	Scaling Vision Transformers to 22 Billion Parameters	Google	United States	Methodology
6	Self-Instruct: Aligning Language Models with Self-Generated Instructions (2023)	Amirkabir University of Technology, Arizona State University, Johns Hopkins University	Iran, United States	Background
7	GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints (2023)	Google, Google Research	—	Methodology
8	YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications	Meituan Inc.	—	—
9	DINOv2: Learning Robust Visual Features without Supervision	Inria, Meta	France	—
10	MiniLLM: Knowledge Distillation of Large Language Models (2024)	Microsoft Research, Tsinghua University	China	—
11	Faster Segment Anything: Towards Lightweight SAM for Mobile Applications (2023)	Kyung Hee University	South Korea	Methodology
12	MetaMath: Bootstrap Your Own Mathematical Questions for Large Language Models (2024)	Hong Kong University of Science and Technology, Hong Kong University of Science & Technology, Huawei	China, Hong Kong, United Kingdom	Background
13	Large Language Models: A Survey	Amazon Inc, CIIRC CTU, Cologne University of Applied Sciences	Czech Republic, Germany	—
14	Kimi-VL Technical Report	—	—	—
15	Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities	Google, Google Inc.	United States	—
16	A Survey on Model Compression for Large Language Models (2024)	Chinese Academy of Sciences, Renmin University of China	China	—
17	Security and Privacy Challenges of Large Language Models: A Survey	Florida International University	United States	—
18	A Comprehensive Survey of Small Language Models in the Era of Large Language Models: Techniques, Enhancements, Applications, Collaboration with LLMs, and Trustworthiness (2025)	Amazon, Industry Research, LinkedIn	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
19	A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development (2023)	East China University of Science and Technology, Fudan University, Institute of Automation, Chinese Academy of Sciences	Australia, China	Background
20	A survey on deep neural network pruning: Taxonomy, comparison, analysis, and recommendations (2024)	Harbin Institute of Technology, Harbin Institute of Technology (Shenzhen), The University of Adelaide	Australia, China	—
21	Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence	Birla Institute of Technology and Science, Birla Institute of Technology and Science (BITS), BITS Pilani	China, India, Italy	**Methodology**
22	A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT	Beihang University, Duke University, Hangzhou Dianzi University	Australia, China, Singapore	—
23	Fast Inference from Transformers via Speculative Decoding (2023)	Google, Google Research	United States	Background
24	CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society	King Abdullah University of Science and Technology	Saudi Arabia	—
25	H2O: Heavy-Hitter Oracle for Efficient Generative Inference of Large Language Models (2023)	Adobe Research, Meta AI, Stanford University	United States	—
26	Depth Anything V2 (2024)	ByteDance, The Chinese University of Hong Kong, The University of Hong Kong	China, Hong Kong	—
27	LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS	The University of Texas at Austin, University of Texas at Austin, Xiamen University	China, United States	**Methodology**
28	Deepfake detection using deep learning methods: A systematic and comprehensive review (2022)	Istanbul Aydın University, Kadir Has University	Turkey	**Methodology**
29	Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook	Hamad Bin Khalifa University, United Arab Emirates University, Weill Cornell Medicine-Qatar	Qatar, United Arab Emirates	—
30	Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence	Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Free University of Bozen-Bolzano, Galala University	Egypt, Italy, South Korea	—

Showing the 30 most-cited of 36 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks

"For fair comparison, we do not use knowledge distillation [19] and neural architecture search [79]."

METHODOLOGY One-step Diffusion with Distribution Matching Distillation

"They frame diffusion distillation as knowledge distillation [19], where a student model is trained to distill the multi-step outputs of the original diffusion model into a single step."

METHODOLOGY Scaling Vision Transformers to 22 Billion Parameters

"We perform model distillation (Hinton et al., 2015) to compress the ViT-22B into smaller, more widely usable ViTs."

METHODOLOGY GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints

"Model distillation (Hinton et al., 2015; Gou et al., 2021) instead reduces model size at a given precision, using data generated from the larger model to finetune the smaller model."

METHODOLOGY Faster Segment Anything: Towards Lightweight SAM for Mobile Applications

"In essence, this retraining process is knowledge distillation Hinton et al. [2015], which transfers the knowledge from a ViT-H-based SAM to a SAM with a smaller image encoder (see Figure 2 left)."

FOLLOW-UP WORK

Deep learning

2015 · Nature · 111,795 citations (GS)

Field-normalised: 74,422 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Rewrite the Stars (2024)	Microsoft, Northeastern University	United States	—
2	Adding Conditional Control to Text-to-Image Diffusion Models	Stanford University	United States	—
3	What do we need to build explainable AI systems for the medical domain? (2017)	Medical University Graz, The University of Manchester, Universität Hamburg	Austria, Cyprus, Germany	Background
4	Activation functions: Comparison of trends in practice and research for deep learning (2018)	University of Strathclyde	United Kingdom	**Methodology**
5	On the Opportunities and Risks of Foundation Models	Stanford Institute for Human-Centered Artificial Intelligence, Stanford University	United States	—
6	Large Language Models for Robotics: A Survey (2023)	University of Illinois Chicago	United States	Background
7	From System 1 to System 2: A Survey of Reasoning Large Language Models (2025)	AiShiWeiLai AI Research, Chinese Academy of Sciences, City University of Hong Kong and the Hong Kong University of Science and Technology (Guangzhou)	China, United Arab Emirates, United Kingdom	—
8	Fake news detection on social media: A data mining perspective (2017)	Arizona State University, Charles River Analytics, Michigan State University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
9	A Survey on Deep Learning: Algorithms, Techniques, and Applications (2018)	—	—	—
10	A Survey of Deep Active Learning (2021)	Carnegie Mellon University, Monash University, National Institute of Technology Kurukshetra	Australia, China, India	—
11	Transformers in Vision: A Survey (2022)	Inception Institute of Artificial Intelligence, MBZ University of Artificial Intelligence, Monash University	Australia, United Arab Emirates, United States	Background
12	Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey (2023)	Amazon AWS AI Labs, Harvard University, Synoptic Engineering	Spain, United States	—
13	A Survey on Evaluation of Large Language Models (2024)	Carnegie Mellon University, Hong Kong University of Science and Technology, Institute of Automation, Chinese Academy of Sciences	China, Hong Kong, United States	Background
14	Deep Multimodal Data Fusion	The University of Alabama at Birmingham	United States	Background
15	Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review (2023)	Ibn Zohr University, University Ibn Zohr	—	Background
16	Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers	Bowie State University, Morgan State University, University of the District of Columbia	United States	Background
17	Deep Learning for Health Informatics (2016)	Imperial College London	United Kingdom	Background
18	Deep convolutional neural network for inverse problems in imaging (2017)	Dassault Aviation, École Polytechnique Fédérale de Lausanne, École polytechnique fédérale de Lausanne (EPFL)	France, Switzerland	Background
19	Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review (2017)	University of South Africa	South Africa	—
20	Efficient Processing of Deep Neural Networks: A Tutorial and Survey (2017)	Massachusetts Institute of Technology	United States	Background
21	Object Detection with Deep Learning: A Review (2019)	Hefei University of Technology, University of Louisiana at Lafayette	China, United States	**Methodology**
22	Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges (2020)	Robert Bosch GmbH, Ulm University, University of Stuttgart	Germany	Background
23	A Survey of the Usages of Deep Learning for Natural Language Processing (2020)	University of Colorado Colorado Springs	United States	Background

No.	Citing paper	Citing institution(s)	Country	S2
24	AI in Medical Imaging Informatics: Current Challenges and Future Directions (2020)	AstraZeneca, Boston Healthcare System, Emory University	Cyprus, Greece, New Zealand	—
25	A Unifying Review of Deep and Shallow Anomaly Detection (2021)	Fraunhofer Heinrich Hertz Institute, Fraunhofer Heinrich Hertz Institute (HHI), Oregon State University	Germany, United States	Background
26	A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises (2021)	Case Western Reserve University, Case Western Reserve University and Louis Stokes Cleveland Veterans Administration Medical Center, Emory University and Georgia Institute of Technology	China, Germany, Israel	Background
27	Self-Supervised Learning: Generative or Contrastive (2021)	Beijing Institute of Technology, Renmin University of China, Tsinghua University	China	—
28	Domain Adaptation for Medical Image Analysis: A Survey (2021)	University of North Carolina at Chapel Hill	United States	—
29	Deep learning for electroencephalogram (EEG) classification tasks: a review (2019)	University of Houston	United States	—
30	Introduction to Machine Learning, Neural Networks, and Deep Learning (2020)	Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Casey Eye Institute, Oregon Health & Science University, Massachusetts General Hospital	United States	—

Showing the 30 most-cited of 1,021 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Activation functions: Comparison of trends in practice and research for deep learning

"The numerous deep architectures include deep feedforward NNs, CNN, long short term memory, RNN, DBN, and deep generative models like deep Boltzmann machines, etc [1], [4]."

METHODOLOGY Object Detection with Deep Learning: A Review

"CNN is the most representative model of deep learning [26]."

Contribution 2

Claim — Contribution 2

The researcher pioneered deep convolutional neural networks for image classification and advanced visual representation learning through contrastive methods and normalization techniques.

The researcher's foundational contribution rests on the 2012 paper 'ImageNet classification with deep convolutional neural networks,' published in NIPS. This work appears to have established a critical benchmark for applying deep learning to large-scale image recognition tasks, serving as the cornerstone for subsequent advancements in the field.

This line of work addresses the challenge of effectively training deep models for visual data. The titles suggest a progression from establishing the core architecture for classification to refining training stability and representation quality. Follow-up works, such as 'Layer Normalization' (2016) and 'A simple framework for contrastive learning of visual representations' (2020), indicate a sustained effort to improve model robustness and self-supervised learning capabilities, building directly upon the initial breakthrough.

The significance of this research is evidenced by its extensive uptake within the scientific community. The core paper has accumulated over 194,000 citations, while the follow-up works have garnered approximately 31,700 and 18,700 citations respectively. Furthermore, analysis of citing papers reveals that 98.4% of citations originate from independent researchers, underscoring the broad, field-wide impact and independent validation of these contributions.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 1,110 · **177 flagged influential by Semantic Scholar**

CORE PAPER

ImageNet classification with deep convolutional neural networks

2012 · Advances in Neural Information Processing Systems 25 (NIPS 2012) · 194,363 citations (GS)

Field-normalised: 128,391 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2012 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Xception: Deep Learning with Depthwise Separable Convolutions (2017)	Google	—	Background
2	Unsupervised Feature Learning via Non-Parametric Instance Discrimination (2018)	The Chinese University of Hong Kong, UC Berkeley	China, United States	Methodology
3	VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection (2018)	Apple Inc	—	Background
4	Designing Network Design Spaces (2020)	Facebook AI Research	—	Background
5	ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks (2020)	Dalian University of Technology, Harbin Institute of Technology, Tianjin University	China	Background
6	A ConvNet for the 2020s (2022)	Facebook, Meta AI, UC Berkeley	United States	Methodology
7	MetaFormer is Actually What You Need for Vision (2022)	Huazhong University of Science and Technology, National University of Singapore, Sea AI Lab	China, Singapore, United States	Methodology
8	HexPlane: A Fast Representation for Dynamic Scenes (2023)	University of Michigan	United States	—
9	Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks (2023)	Hong Kong University of Science and Technology, Rutgers University, Texas State University	Hong Kong, United States	Background
10	EVA: Exploring the Limits of Masked Visual Representation Learning at Scale (2023)	Beijing Academy of Artificial Intelligence, Beijing Institute of Technology, Huazhong University of Science and Technology	China	Background

No.	Citing paper	Citing institution(s)	Country	S2
11	SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy (2023)	East China Normal University, Tongji University	China	—
12	EfficientViT: Memory Efficient Vision Transformer With Cascaded Group Attention	Microsoft Research, The Chinese University of Hong Kong	Hong Kong	Methodology
13	InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions (2023)	Nanjing University, SenseTime, SenseTime Research	China, Hong Kong	Methodology
14	VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking (2023)	Nanjing University, Shanghai AI Lab, Shanghai Artificial Intelligence Laboratory	China	Background
15	ConvNeXt V2: Co-Designing and Scaling ConvNets With Masked Autoencoders	KAIST, Meta AI, New York University	South Korea, United States	—
16	InternVL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks	Nanjing University, SenseTime, SenseTime, Shanghai AI Laboratory	China, Hong Kong	Background
17	UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition (2024)	Tencent, The Chinese University of Hong Kong	China, Hong Kong	Methodology
18	Rewrite the Stars	Microsoft, Northeastern University	United States	—
19	EMCAD: Efficient Multi-scale Convolutional Attention Decoding for Medical Image Segmentation	The University of Texas at Austin	United States	Methodology
20	TransNeXt: Robust Foveal Visual Perception for Vision Transformers (2023)	Independent Researcher	—	—
21	Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks	Microsoft	United States	Methodology
22	Boosting Continual Learning of Vision-Language Models via Mixture-of-Experts Adapters (2024)	Dalian University of Technology, Tsinghua University, University of Electronic Science and Technology of China	China	Methodology
23	MambaVision: A Hybrid Mamba-Transformer Vision Backbone (2024)	NVIDIA	United States	—
24	SlowFast Networks for Video Recognition (2019)	Facebook	—	Background
25	Multiscale Vision Transformers (2021)	Facebook, Meta AI	—	Background
26	Scalable Diffusion Models with Transformers	—	—	Methodology
27	Cross-Entropy Loss Functions: Theoretical Analysis and Applications (2023)	Courant Institute of Mathematical Sciences, Google Research, New York University	United States	Background

No.	Citing paper	Citing institution(s)	Country	S2
28	[MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications](#) (2017)	Google Inc.	—	**Influential**
29	[Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour](#) (2017)	Facebook AI Research	United States	Background
30	[Rethinking Atrous Convolution for Semantic Image Segmentation](#) (2017)	Google Inc.	United States	**Methodology**

Showing the 30 most-cited of 1,039 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** Unsupervised Feature Learning via Non-Parametric Instance Discrimination
>
> *"When fine-tuning AlexNet and VGG16, we follow the standard practice, fixing the conv1 model weights."*

> **METHODOLOGY** A ConvNet for the 2020s
>
> *"The introduction of AlexNet [40] precipitated the "ImageNet moment" [59], ushering in a new era of computer vision."*

> **METHODOLOGY** MetaFormer is Actually What You Need for Vision
>
> *"Raghu et al. [42] compare the feature difference between ViT and CNNs, finding that self-attention enables early aggregation of global information while residual connections strongly propagate features from lower to higher layers."*

> **METHODOLOGY** EfficientViT: Memory Efficient Vision Transformer With Cascaded Group Attention
>
> *"However, most of them target at minimizing Flops and parameters [16], which could have low correlations with actual inference latency [70] and still inferior to efficient CNNs in speed."*

> **METHODOLOGY** InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions
>
> *"Straining from AlexNet [32], lots of deeper and more effective neural network architectures have been proposed, such as VGG [33], GoogleNet [34], ResNet [35], ResNeXt [36], EfficientNet [37, 38], etc."*

FOLLOW-UP WORK

[A simple framework for contrastive learning of visual representations](#)

2020 · International conference on machine learning, 1597-1607, 2020 · 31,749 citations (GS)

Field-normalised: 23,922 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2020 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	[ImageBind: One Embedding Space To Bind Them All](#) (2023)	Meta AI	—	**Result**
2	[VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking](#) (2023)	Nanjing University, Shanghai AI Lab, Shanghai Artificial Intelligence Laboratory	China	—
3	[SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery](#) (2024)	Ant Group, Ant Group / MYBank, National University of Singapore	China, Singapore, United States	—
4	[Improved Baselines with Visual Instruction Tuning](#) (2024)	Microsoft Research, University of Wisconsin–Madison	United States	**Methodology**
5	[Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks](#)	Microsoft	United States	**Methodology**

No.	Citing paper	Citing institution(s)	Country	S2
6	EfficientSAM: Leveraged Masked Image Pre-training for Efficient Segment Anything	Meta	—	Background
7	SimCSE: Simple Contrastive Learning of Sentence Embeddings (2021)	Princeton University, Tsinghua University	China, United States	Methodology
8	BEiT: BERT Pre-Training of Image Transformers	Harbin Institute of Technology, Microsoft Research	China	—
9	On the Opportunities and Risks of Foundation Models (2021)	Stanford Institute for Human-Centered Artificial Intelligence, Stanford University	United States	—
10	DINOv2: Learning Robust Visual Features without Supervision	Inria, Meta	France	Methodology
11	Contrastive Preference Optimization: Pushing the Boundaries of LLM Performance in Machine Translation (2024)	Johns Hopkins University, Microsoft	United States	Background
12	Representation Alignment for Generation: Training Diffusion Transformers Is Easier Than You Think	—	—	—
13	Expanding Performance Boundaries of Open-Source Multimodal Models with Model, Data, and Test-Time Scaling (2024)	Fudan University, Nanjing University, SenseTime Research	China	—
14	Diffusion Models in Vision: A Survey	University of Bucharest, University of Central Florida	Romania, United States	—
15	SpectralGPT: Spectral Remote Sensing Foundation Model	Aerospace Information Research Institute, Aerospace Information Research Institute, Chinese Academy of Sciences, Helmholtz-Zentrum Dresden-Rossendorf	Australia, China, France	Background
16	RemoteCLIP: A Vision Language Foundation Model for Remote Sensing (2024)	Chinese Academy of Forestry, Griffith University, Hohai University	Australia, China	—
17	The Faiss Library (2025)	DeepSeek, FAIR, Meta, Kyutai	China, France, United States	—
18	Foundation models in robotics: Applications, challenges, and the future	Physical Intelligence, Princeton University, Rhoda AI	Canada, China, United States	—
19	Learning to Prompt for Vision-Language Models	Hong Kong Baptist University, Nanyang Technological University	China, Singapore	Background
20	A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT	Beihang University, Duke University, Hangzhou Dianzi University	Australia, China, Singapore	—
21	Masked Autoencoders Are Scalable Vision Learners (2022)	Facebook	—	Influential
22	Emerging Properties in Self-Supervised Vision Transformers (2021)	Facebook AI Research, Inria	France, United States	Methodology

No.	Citing paper	Citing institution(s)	Country	S2
23	Multimodal Foundation Models: From Specialists to General-Purpose Assistants (2024)	Microsoft Research	—	—
24	A foundation model for generalizable disease detection from retinal images	NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, University College London, University of Coruña	Spain, United Kingdom, United States	Methodology
25	A whole-slide foundation model for digital pathology from real-world data (2024)	Microsoft Research, Providence Genomics, Providence Health System	United States	—
26	Towards a general-purpose foundation model for computational pathology	Brigham and Women's Hospital, Brigham and Women's Hospital, Harvard Medical School, Brigham and Women's Hospital, Harvard Medical School	United States	—
27	Self-supervised learning for medical image classification: a systematic review and implementation guidelines (2023)	Stanford University	United States	—
28	A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution.	Allen Institute for Brain Science, Boston College, Google Research	Switzerland, United States	—
29	Foundation models defining a new era in vision: a survey and outlook (2025)	Khalifa University, MBZ University of AI	United Arab Emirates	—
30	Conditional Prompt Learning for Vision-Language Models (2022)	Nanyang Technological University	Singapore	—

Showing the 30 most-cited of 37 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology* / *Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

RESULT ImageBind: One Embedding Space To Bind Them All

"This is in contrast to standard self-supervised methods like SimCLR [10] whose performance improves with MLP projection heads."

METHODOLOGY Improved Baselines with Visual Instruction Tuning

"Inspired by the improved performance in self-supervised learning by changing from a linear projection to an MLP [9, 10], we find that improving the vision-language connector's representation power with a two-layer MLP can improve LLaVA's multimodal capabilities, compared with the original linear..."

METHODOLOGY Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks

"In pursuit of a versatile vision foundation model, we re-visit three predominant pre-training paradigms: supervised (e.g ., ImageNet classification [18]), self-supervised (e.g ., SimCLR [9], MoCo [25], BEiT [4], MAE [24]), and weakly supervised (e.g ., CLIP [64], Florence [95], SAM [32])."

METHODOLOGY SimCSE: Simple Contrastive Learning of Sentence Embeddings

"We follow the contrastive framework in Chen et al. (2020) and take a cross-entropy objective with in-batch negatives (Chen et al., 2017; Henderson et al., 2017): let hi and h+i denote the representations of xi and x+i , the training objective for (xi, x+i) with a mini-batch of N pairs is: `i = −..."

METHODOLOGY DINOv2: Learning Robust Visual Features without Supervision

"For our semantic segmentation evaluation, we consider two different setups."

FOLLOW-UP WORK

Layer Normalization

2016 · arXiv preprint · 18,702 citations (GS)

Field-normalised: 12,348 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2016 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks (2023)	Hong Kong University of Science and Technology, Rutgers University, Texas State University	Hong Kong, United States	—
2	EVA: Exploring the Limits of Masked Visual Representation Learning at Scale	Beijing Academy of Artificial Intelligence, Beijing Institute of Technology, Huazhong University of Science and Technology	China	**Methodology**
3	UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition	Tencent, The Chinese University of Hong Kong	China, Hong Kong	**Methodology**
4	mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration	—	—	—
5	Molmo and PixMo: Open Weights and Open Data for State-of-the-Art Vision-Language Models	Allen Institute for AI, University of Pennsylvania, University of Washington	United States	—
6	Segment Anything	Meta	—	Background
7	Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference (2024)	Answer.AI, Hugging Face, Johns Hopkins University	France, United States	—
8	A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT (2023)	Lehigh University, University of Illinois at Chicago	United States	**Methodology**
9	Retentive Network: A Successor to Transformer for Large Language Models	—	—	**Methodology**
10	iTransformer: Inverted Transformers Are Effective for Time Series Forecasting (2024)	Ant Group, Tsinghua University	China	Background
11	Mamba: Linear-Time Sequence Modeling with Selective State Spaces	Carnegie Mellon University, Princeton University	United States	—
12	U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation	University Health Network	Canada	**Methodology**
13	Expanding Performance Boundaries of Open-Source Multimodal Models with Model, Data, and Test-Time Scaling	Fudan University, Nanjing University, SenseTime Research	China	—
14	Efficient Memory Management for Large Language Model Serving with PagedAttention	Independent Researcher, UC Berkeley, UC San Diego	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
15	A Comprehensive Overview of Large Language Models (2024)	Australian National University, King Fahd University of Petroleum and Minerals, The Chinese University of Hong Kong	Australia, China, Pakistan	—
16	A Survey on Large Language Models for Code Generation	NAVER Cloud, The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou)	China, South Korea	—
17	VM-UNet: Vision Mamba UNet for Medical Image Segmentation	Shanghai Jiao Tong University	China	—
18	nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer	The Chinese University of Hong Kong (Shenzhen), The University of Hong Kong, Xiamen University	China, Hong Kong	Methodology
19	Medical image segmentation review: The success of u-net (2024)	Independent Researcher, Mashhad University of Medical Sciences, Mila - Quebec AI Institute	Canada, Germany, Iran	Methodology
20	A Tutorial on Fluid Antenna System for 6G Networks: Encompassing Communication Theory, Optimization Methods and Hardware Designs	The Hong Kong University of Science and Technology	China	Methodology
21	Visual Prompt Tuning (2022)	Cornell University, Meta AI, University of Copenhagen	Denmark, United States	Background
22	MobileNetV4: Universal Models for the Mobile Ecosystem	Google	United States	—
23	Masked Autoencoders Are Scalable Vision Learners (2022)	Facebook	—	Background
24	A Survey of Large Language Models	Renmin University of China, Université de Montréal	Canada, China	Methodology
25	Autoregressive Image Generation without Vector Quantization (2024)	Google DeepMind, Massachusetts Institute of Technology, MIT	China, United Kingdom, United States	Methodology
26	xLSTM: Extended Long Short-Term Memory (2024)	Johannes Kepler University Linz, NXAI	Austria	—
27	YOLOv10: Real-Time End-to-End Object Detection	Tsinghua University, University of Sheffield	China, United Kingdom	Methodology
28	Segment anything in medical images	New York University, University Health Network, Western University	Canada, United States	—
29	Barren plateaus in variational quantum computing (2025)	California Institute of Technology, École polytechnique fédérale de Lausanne (EPFL), Google Quantum AI	Russia, Singapore, Switzerland	—

No.	Citing paper	Citing institution(s)	Country	S2
30	Simple and Controllable Music Generation (2023)	Meta AI	—	Methodology

Showing the 30 most-cited of 34 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology* / *Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY EVA: Exploring the Limits of Masked Visual Representation Learning at Scale

"The output feature of EVA is first normalized [3] and then projected to the same dimension as the CLIP feature via a linear layer."

METHODOLOGY UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition

"A stage comprises blocks whose vanilla design resembles ConvNeXt, i.e ., a depthwise (DW) conv layer and a Feed-Forward Network (FFN) with GRN unit [85], but we use BN instead of LayerNorm [1] after the conv layer as BN can be equivalently merged into the conv layer to eliminate its inference costs."

METHODOLOGY A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT

"Gopher [39] uses a GPT-like structure but replace LayerNorm [63] with RSNorm, where a residual connection is added to the original layernorm structure to maintain the information."

METHODOLOGY Retentive Network: A Successor to Transformer for Large Language Models

"Retentive network (RetNet) is stacked with L identical blocks, which follows a similar layout (i.e., residual connection, and pre-LayerNorm) as in Transformer (Vaswani et al., 2017)."

METHODOLOGY U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation

"After passing the Layer Normalization [2], the features enter the Mamba block that contains two parallel branches."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Stanford University	United States	SCImago #18 · THE =5 · QS 3	181
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	157
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	119
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	116
UC Berkeley	United States	—	95
Google	United States	—	94
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	88
Google Research	United States	—	80
Nanyang Technological University	Singapore	SCImago #137	76
University of Oxford	United Kingdom	SCImago #26 · THE 1 · QS 4	73
New York University	United States	SCImago #116 · THE =31 · QS 55	71
University of Toronto	Canada	SCImago #39 · THE 21 · QS 29	70
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	68
Columbia University	United States	SCImago #65 · THE 20 · QS =38	67
Peking University	China	SCImago #11 · THE 13 · QS 14	65

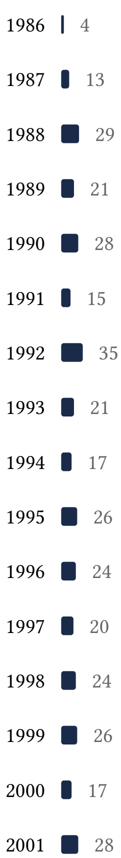
Geographic distribution of citing authors

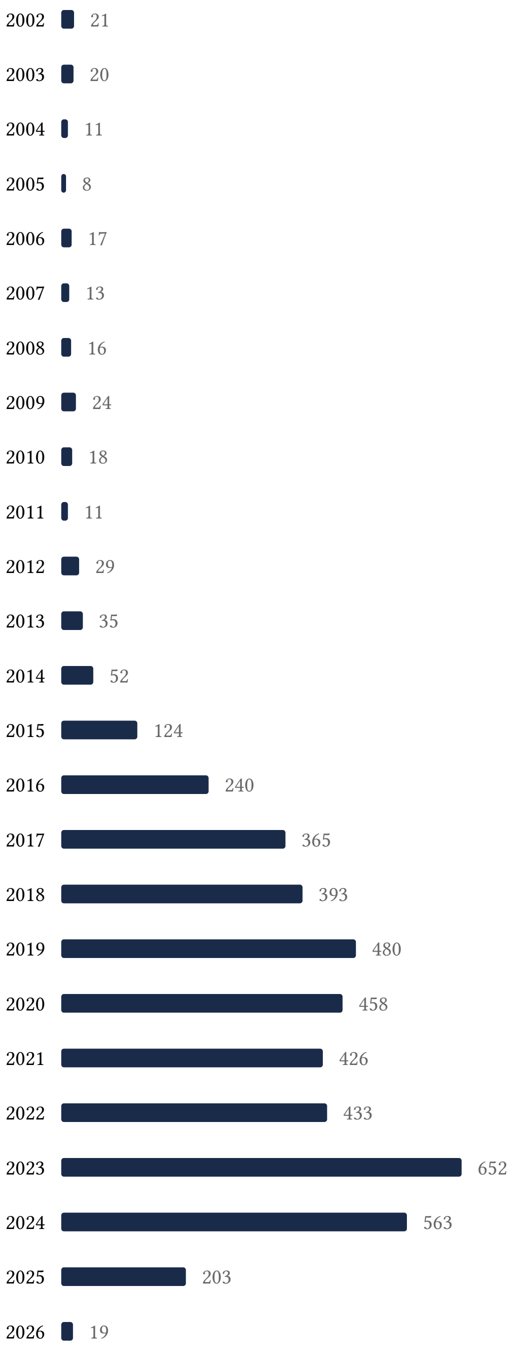
Country	Citing papers
United States	2,206
China	1,111
United Kingdom	535
Germany	361
Canada	306
Australia	255
Singapore	184
Switzerland	162
South Korea	130
France	129
India	114
Italy	112

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.





F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- [ ] Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).

- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition’s merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition’s exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Dropout: A Simple Way to Prevent Neural Networks from Overfitting	1,860	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	ImageNet classification with deep convolutional neural networks	1,110	8 CFR 204.5(h)(3)(v) — Criterion 5