

# Abhinav JAIN

## Research Engineer

in [linkedin.com/in/abhinav-jain-483975aa](https://www.linkedin.com/in/abhinav-jain-483975aa)

✉ [jabhinaviitk@gmail.com](mailto:jabhinaviitk@gmail.com)

☎ +91 951 918 9181 @ USER ID : [abhinavj@in.ibm.com](mailto:abhinavj@in.ibm.com)

📍 Delhi, India

## EDUCATION

APR 2017	<b>INDIAN INSTITUTE OF TECHNOLOGY, KANPUR, INDIA</b>
APR 2013	BACHELOR OF TECHNOLOGY (B.TECH) IN ELECTRICAL ENGINEERING Cumulative Performance Index, CPI : <b>9.0/10.0</b>
APR 2013	<b>JAWAHARLAL NEHRU SCHOOL, BHOPAL, INDIA</b>
APR 2012	ALL INDIA SENIOR SCHOOL CERTIFICATE EXAMINATION, AISSC, CLASS XII Percentage : <b>91.8%</b>

## PROFESSIONAL EXPERIENCE

PRESENT JUL 2017	<b>SOFTWARE DEVELOPER   IBM RESEARCH LABS, INDIA</b> <b>SCANNED PDF-TO-HTML CONVERSION, IBM WATSON SERVICES</b> <i>Project Manager : DR. SAMEEP MEHTA</i> <i>Description :</i> Convert scanned PDF documents into “consumable” representations (HTML/JSON) that can be served as inputs to Business Automation Tool to drive structured data extraction from client documents. Provided support to the PDF-to-HTML pipeline by adding following capabilities : <ul style="list-style-type: none"><li>➤ Hybrid PDF support to extract data from PDFs containing scanned text and programmatic content.</li><li>➤ Reorientation of skewed text for reliable OCR based text extraction.</li><li>➤ Service for rendering document images when language of the text is changed</li><li>➤ Devised ensemble of unsupervised techniques based on connected component analysis for detecting logos, bar-codes and signatures for downstream processing such as querying, retrieval.</li></ul> <div>Java IntelliJ Idea Maven Apache Gradle OpenCV</div>
PRESENT JUL 2019	<b>RESEARCH ENGINEER   IBM RESEARCH LABS, INDIA</b> <b>PROGRAMMING BY EXAMPLE, AI FOR DATA</b> <i>Project Manager : HIMA PATEL</i> <i>Description :</i> With minimum number of user queries, generate annotations for a complete dataset. The designed PbE system first identifies heterogeneous clusters in the input dataset and then selects representative samples for getting annotations from the user. Data transformation program is then learned which is refined in an iterative fashion using user feedback (syntax-based or natural language based). <div>Inductive Program Synthesis Few Shot Learning Java Python</div>
AUG 2016 MAY 2016	<b>INTERN   IBM RESEARCH LABS, INDIA</b> <b>COHERENT VISUAL DESCRIPTION OF TEXTUAL INSTRUCTIONS</b> <i>Project Manager : DR. SAMEEP MEHTA</i> <i>Description :</i> In this project, a multi-stage framework was developed to provide visual aid for a sequence of text based instructions. The visual aid is in the form of coherent images associated with each instruction. The framework works in the following sequential manner : <ul style="list-style-type: none"><li>➤ For each instruction, visualisable phrases are mined using co-reference resolution, POS tagging and Dependency parsing. These phrases consist of head action verbs and associated noun phrases.</li><li>➤ For each visualisable phrase, an API query is made to retrieve a set of images from the dataset crawled from sources such as WikiHow, Flickr, Google etc.</li><li>➤ To maintain coherency across multiple instructions sharing common information in the form of latent/non-latent entities, a graph based matching method utilising Dijkstra’s algorithm is proposed.</li></ul> We conducted a user study to validate (1) improvement in understanding of text instructions and (2) resemblance to actual ground truth. <a href="https://github.com/jabhinav/IBM-project">github.com/jabhinav/IBM-project</a> <div>Python WikiHow textToImage text mining Graph Matching</div>

### VIDEO REPRESENTATION LEARNING FOR FINE-GRAINED SCENE RECOGNITION AND RETRIEVAL

2018-19

[github.com/jabhinav/Deep-Video-Understanding](https://github.com/jabhinav/Deep-Video-Understanding)

Advertising in digital media often requires recognition of critical scenes in videos for smart placement of brand advertisements. These critical scenes raise viewer anxiety and are a part of some parent activity. We distinguish them from the rest of non-critical scenes using an order-preserving fine-grained similarity metric that learns the required representations. The learned metric is tested in two novel tasks : video critical **scene recognition** and fine-grained **video retrieval**. To learn the metric, we proposed Pentuplet Loss [3] and later on, an improved and more robust Radial Loss [4]. These losses exploit 'Quadlet Sampling' to mine data where each training sample is a tuple of query, positive, intermediate and negative samples. Lastly, to ascertain the effectiveness of the loss in learning a deep metric for measuring similarities, we tested its performance against state-of-the-art baselines in the known tasks : fine-grained image retrieval and shot-boundary detection.

Deep Metric Learning Event Recognition Content Based Retrieval or Ranking Shot Boundary Detection CNN-LSTM Siamese Nets Loss Formulation

### DEEP LEARNING FOR DOCUMENT IMAGE QUALITY ENHANCEMENT

2017-19

[github.ibm.com/abhinavj/Super-Resolving-Documents](https://github.com/abhinavj/Super-Resolving-Documents)

Poor quality scanned document images suffer from low token fidelity when an OCR engine such as Tesseract is used for token extraction. To remedy this, we leveraged deep learning based solutions for document quality enhancement ([6] and [7]) and delivered the same for public release as part of IBM's Watson API. In [6], we formulated a novel 'Text Quality Improvement Loss' for the standard super-resolution convolutional neural network (SRCNN) to generate high-resolution text images. The proposed framework identifies text regions from images and minimizes additional MSE between such localised regions on top of the standard MSE, enforced by Single Image Super Resolution frameworks. This results in simultaneous optimisation of perceptual quality of the image and the OCR performance.

SuperResolution Caffe Tesseract SRCNN OCR Boosting UNLV Dataset Testing

### EDUCATION ENRICHMENT

2017-18

[github.com/jabhinav/EducationEnrichment](https://github.com/jabhinav/EducationEnrichment)

Formal texts such as journal articles are composed of complex terminologies intended to be understood by targeted demographic. In absence of domain knowledge, they tend to be more ambiguous for general readers. To avail a complete semantic understanding of such texts for the readers, we proposed an enrichment system [2] that mitigates the problem of searching for required information through heaps of sources. The system augments given text with required concept definitions, applications and concept dependency graphs. Our framework extracts key-concepts (technical terms) based on user discretion via a sequence of filtering stages - Linguistic Filtering, BBC Pruning and StackExchange Pruning. It detects the presence of required information by classifying each associated sentence into definition/application of the key-concept using a CNN-LSTM network. The same framework also runs on a data source such as Wikipedia to return the concept's missing definition or real-life application.

NLP AI for Education Sentence Classification Information Extraction Dependency Parsing

### EVOLVING AI

2017-18

[github.com/jabhinav/Model-Learning](https://github.com/jabhinav/Model-Learning)

In this project, we address the problem of re-training a deep neural network for a new class with limited training data ('n' to 'n+1' class learning) using a novel concept of Deep part embeddings (DPEs). DPEs are sub-networks of neuron activation extracted from a trained network identifying a visual and distinguishable element of a class. We identify visual elements that intuitively constitute a new class and extract the corresponding DPEs from the network pre-trained for the class sharing the identified visual element. Finally, we assemble them into a new network and re-train the model on limited samples of the new class and a subset of data from 'n' classes to achieve high accuracy on the new class without significantly affecting the accuracy of n classes. We studied and produced results for DPE integration under two configurations : (i) sequential, when DPEs are sourced from different CNN architectures and (ii) shared; when DPEs are sourced from the same CNN architecture.

Few-shot learning Knowledge Transfer Model Learning Activation Filtering Convolutional Neural Networks Deep Part Embeddings

## PUBLICATIONS

- |          |   |
|----------|---|
| DEC 2017 | [1] S Mujumdar, N Gupta, <u>A Jain</u> , S Mehta, "Coherent Visual Description of Textual Instructions," in <i>IEEE International Symposium on Multimedia (ISM)</i> .   |
| AUG 2018 | [2] <u>A Jain</u> , N Gupta, S Mujumdar, S Mehta, R Madhok, "Content Driven Enrichment of Formal Text using Concept Definitions and Applications," in <i>Proceedings of the 29th on Hypertext and Social Media (HT)</i> . |
| AUG 2018 | [3] N Gupta, <u>A Jain</u> , P Agarwal, S Mujumdar, S Mehta, "Pentuplet Loss for Simultaneous Shots and Critical Points Detection in a Video," in <i>International Conference on Pattern Recognition (ICPR)</i> .         |

- APR 2019 | [4] A Jain, P Agarwal, S Mujumdar, N Gupta, S Mehta, C Chattopadhyay, "Radial Loss for Learning Fine-grained Video Similarity Metric," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- APR 2019 | [5] N Gupta, S Mujumdar, P Agarwal, A Jain, S Mehta, "Learning Convolutional Neural Networks with Deep Part Embeddings," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- SEP 2019 | [6] S Mujumdar, N Gupta, A Jain, D. Burdick, "Simultaneous Optimisation of Image Quality Improvement and Text Content Extraction from Scanned Documents," in *IEEE International Conference on Document Analysis and Recognition (ICDAR)*.
- [7] S Mujumdar, N Gupta, A Jain, D. Burdick, "Multi-Stage Framework to Boost OCR performance on Low Quality Document Images," (Under Submission)

## PATENTS

- 2018 | **System and Method to Generate Dynamic Personalized Infographics**  
A system is proposed to automatically generate info-graphics from floating social media trends, personalised to user interests with the help of unstructured data available across different online platforms like Facebook, Twitter, News etc. The generated info-graphics are personalized by specifying the visual and non-visual elements specific to the user such as sketch or caricature view for the former and text based personalisation elements such as viewer age-group, emotions, sentiment, sarcasm etc. for the latter.

## ACADEMIC PROJECTS

- Nov 2016 | **TOPIC MODELLING**  
Aug 2016 | *Supervisor* : Dr. Gaurav Pandey, Dr. Piyush Rai, I.I.T KANPUR  
*Course* : UNDER-GRADUATE PROJECT
- Discover underlying latent themes(topics, meta-topics) of Newsgroup Corpora using topic modelling.
  - Formulated a survey report comparing hitherto proposed state-of-the-art algorithms - Poisson Factor analysis (PFA), PFA with Sigmoid Belief Networks, Deep PFA (DPFA) and Discriminative-DPFA.
  - Compared aforementioned algorithms based on their formulation, underlying assumptions, advantages-disadvantages and their document label classification performance.
- Topic Modelling | Document classification | Probabilistic Machine Learning | Gibbs Sampling
- Jul 2015 | **VISION BASED SURVEILLANCE AND TRACKING SYSTEM FOR UAVS**  
May 2015 | *Supervisor* : Dr. N.K. Verma, I.I.T KANPUR  
*Course* : SUMMER PROJECT
- Real-time implementation of the following vision based object tracking algorithms to gain hands-on experience - Optical Flow Approach, Background Subtraction, Interest Point Tracking and Real-Time Compressive Tracking.
  - Extensive testing and comparison of the algorithms' robustness against the following factors - pose variation, illumination, occlusion, and motion blur.
  - Integrated the best performing, Real-Time Compressive Tracking with UAV for real-time testing.
- Unsupervised Object Tracking | Survey | Computer Vision
- Dec 2016 | **BAYESIAN OPTIMIZATION FOR HYPER-PARAMETER TUNING**  
Aug 2016 | *Supervisor* : Dr. Purushottam Kar, I.I.T KANPUR  
*Course* : OPTIMIZATION TECHNIQUES
- Studied and compared Bayesian Optimization frameworks - Gaussian Process based and Deep Network for Global Optimization, DNGO.
  - Demonstrated the improvements offered by the Bayesian optimization selection strategies against random selection and hard coding done by experts for automatic hyper-parameter tuning in the task of MNIST digit classification via sparse auto-encoders stacked with softmax classifier.
- Bayesian Optimisation | Gaussian Process | DNGO | Spearmint | Matlab-Python

## FUTURE COURSEWORK

---

Logic And AI  
Topics in Deep Learning  
Learning based Methods in Computer Vision  
Designing AI to cultivate Human Well-Being  
Representation Learning in Computer Vision

## COMPLETED COURSEWORK - IIT KANPUR

---

Data Structures and Algorithms	Probability and Statistics
Probabilistic Machine Learning	Fundamentals of Computing
Introduction to Stochastic Processes	Image Processing
Partial Differential Equation	Visual Recognition
Online Learning and Optimization	Linear Algebra

## EXTRACURRICULAR ACTIVITIES - IIT KANPUR

---

2015	Event Coordinator, Antaragni Leadership Initiative(ALI)
2013-14	Runner-Up, Football Freshers Inferno
2014-15	Secretary, Fine Arts Club
2013-15	Member, Formula SAE

## SKILLS

---

Java	● ● ● ○
Python	● ● ● ●
Caffe	● ● ● ○
Keras	● ● ● ●
OpenCV	● ● ● ●

## AWARDS AND ACHIEVEMENTS

---

2014-15	<b>ACADEMIC EXCELLENCE</b> Awarded for exceptional performance in academics at I.I.T Kanpur
JUL 2016	<b>BEST POSTER AWARD, IBM RESEARCH LABS</b> Achieved recognition for outstanding presentation carried out during internship
2013-14	<b>RUNNERS UP - FOOTBALL, FRESHERS INFERNO</b> Awarded Runners Up trophy at the annual tournament for freshers
2012-13	<b>BEST ALL ROUNDER AWARD</b> Awarded Student of the year in the senior year of high school for all-round excellence