Whitepaper: Using AI to Enhance the Quality of Retail Product Metadata

Technical Report · March 2018

1 author:

Rajkumar Joseph
Cognizant Technology Solutions, chennai, india
8 PUBLICATIONS   **11** CITATIONS

Some of the authors of this publication are also working on these related projects:

Retail metadata extraction View project

Diabetic Retinopathy Detection View project

DIGITAL SYSTEMS & TECHNOLOGY

Using AI to Enhance the Quality of Retail Product Metadata

By increasing the transparency of product information metadata, retailers can help consumers make more informed purchase decisions – and compete more effectively with digital pure-plays. Here's how retailers can accomplish this goal, using machine learning and deep learning techniques.

March 2018

EXECUTIVE SUMMARY

With online sales growing faster than ever,[1] traditional retailers are increasing their investments in omnichannel strategies and redoubling their efforts to meet online consumer demands. One of the most effective ways to keep pace with the giants of e-commerce is to offer superior product discovery and selection capabilities, which requires detailed product information and critical product-specific attributes, coupled with semantic search.

To enhance online product discovery, retailers must maintain and provide digital images and videos, catalog descriptions, category-specific metadata (e.g., nutrition information for food products), stock availability, product matrices (e.g., size ranges), company/brand logos, product ratings and reviews, pricing, and promotions information for all physical stock keeping units (SKU). Acquiring this information from suppliers is a time-consuming task, requiring various methods and a significant amount of manual activity.

Concurrently, many retailers face tremendous product data management challenges as product data is stored in different locations and formats. Another challenge is duplicate data. As a result, many retailers have incomplete and inaccurate product information on their websites and in their systems, with little adherence to data standards and controls, which undermines their competitiveness.

To alleviate this problem, we have built a system that extracts product attributes from food product label images, using computer vision, natural language processing (NLP), optical code recognition (OCR) and machine learning/deep learning techniques. Using these technologies, the system can extract product metadata such as product title, product description, volume/weight, nutrition facts, company/product logos and barcode.

Test results in our labs show 95% accuracy for attribute extraction from high-quality product images featuring machine-printed characters with contrasting backgrounds. This white paper offers perspective on how retailers can take advantage of this solution to get their product metadata house in order.

THE PRODUCT METADATA DILEMMA

In an increasingly e-commerce-driven retail environment, retailers need quality metadata and powerful search platforms to entice customers and help them make effective purchase decisions. However, their inability to easily deliver complete and accurate product information puts them at a severe disadvantage.

Retailers typically rely on suppliers to provide product images and metadata through various methods (electronic data interchange, printed or digital catalogs) and various formats (text, Excel, PDF, XML). Different suppliers often supply inconsistent content for the same products, and few share usable images. Retailers also don't have a simple way of validating the metadata and images before storing the content on their respective systems.

Further, retailers often purchase product information from third-party providers and online Universal Product Code (UPC) databases. UPC product information is used as an input to databases that validate available product metadata. However, online metadata databases are not always accurate; in fact, the data sometimes differs from one UPC database to another.

And yet, next to price, high-quality product images and product metadata (nutrition/ingredients, any special warning messages about the product) are a primary driver for consumer purchases. According to Retailer Brand Services, 97.7% of shoppers expect retailers to show comprehensive product data.[2] In fact, there is an indisputable link between the quality and completeness of online product content and sales (see Figure 1).

Product Information Quality and Completeness Impacts Product Sales
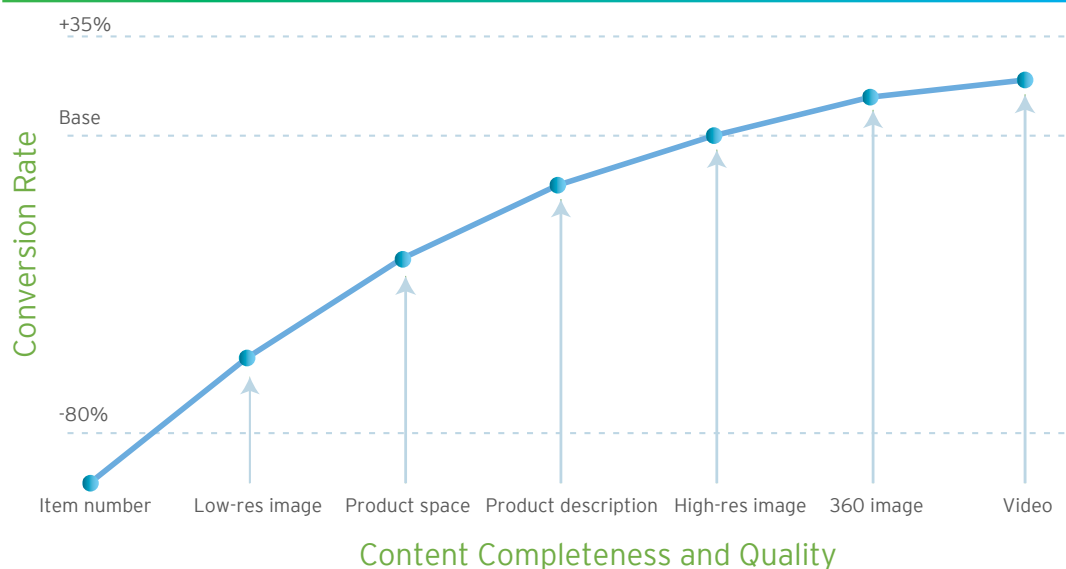
Source: Shotfarm, 2016
Figure 1

Retailers also face several technology challenges when trying to extract content from retail product label images, including region segmentation, diverse product backgrounds, natural settings, typographic and font usage, cursive/handwritten text, lighting conditions, camera artifacts and low-quality images. Other obstacles include:

- **Product size:** Variations in product size limit the product details that the camera can capture and determine the camera that should be used.

- **Dispersed information:** Product metadata may either exist on the external packaging (top, bottom or side) or on the product itself, which can only be seen when the product is unpacked.

- **Information alignment:** Text may be aligned at different angles, posing a challenge for easy extraction.

THE IMPACT OF POOR QUALITY DATA

In the ever-evolving online marketplace, there is still much to be studied and discovered about which types of content best influence shopper purchasing decisions. There are no definitive rules to be followed regarding the optimal number of product images or videos to be displayed by category or sector, or about the preferred character length of product descriptions. However, potential complexities notwithstanding, a solid formula for online success is:

Incremental improvement in the accuracy and completeness of online product content = incremental increase in sales performance

Inaccurate or incomplete product content is damaging to both shoppers' perception of and trust in brands and retailers. In fact, a 2015 study calculated that consumers return $642.6 billion in goods each year, or an estimated 4.4% of $14.5 trillion in global retail sales.[3] While the biggest reason for returns is poor-quality products or the purchase of the wrong item, a big reason for e-commerce returns specifically is that the item didn't match the description (see Figure 2).

Product Information Influences E-Commerce Returns
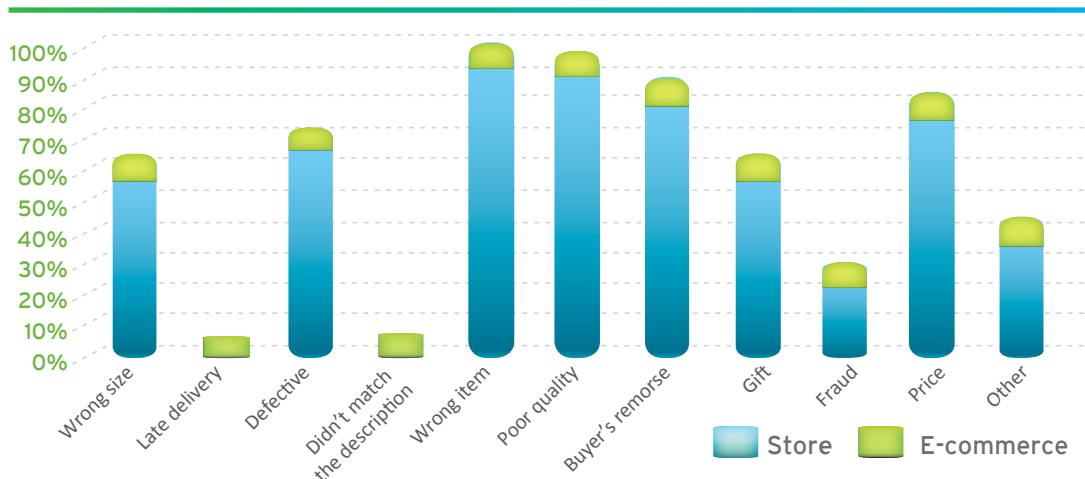
Source: IHL Group/Order Dynamics
Figure 2

While the biggest reason for returns is poor-quality products or the purchase of the wrong item, a big reason for e-commerce returns specifically is that the item didn't match the description

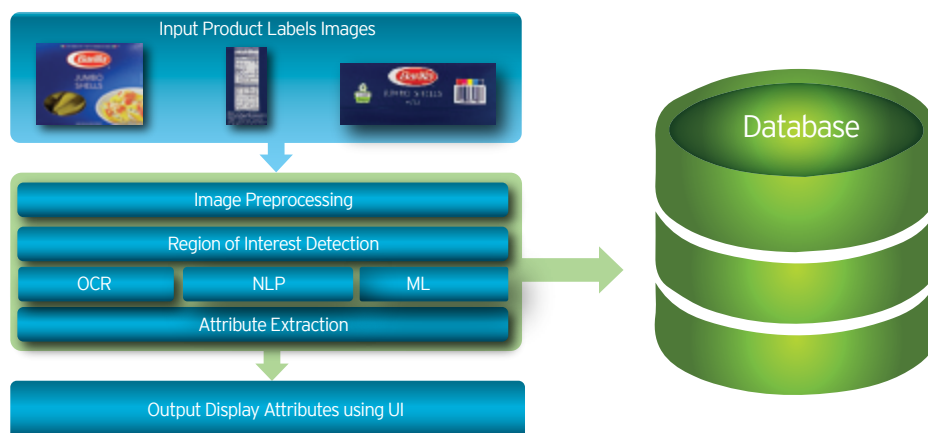## A Conceptual Architecture for Extracting Metadata from Retail Product Images

Figure 3

EFFECTIVE EXTRACTION OF PRODUCT METADATA

We have built a solution to ease the extraction of retail product metadata from product label images. As depicted in Figure 3, our solution photographs a retail product carton from all sides. The captured images are fed into an algorithm that performs data extraction. Image pre-processing techniques are then applied to identify various regions of interest.

For each of these regions of interest, the text attributes are extracted using OCR, and are then improved using machine learning and natural language processing (NLP) techniques before being saved to a database. Similarly, brand and food certification logo detection is conducted using computer vision and machine learning techniques. The attributes extracted include brand name, product name, logo, food certification logo, net weight, nutrition facts and bar code. (See Figure 4 , next page, for a visual summary of these extraction steps.)
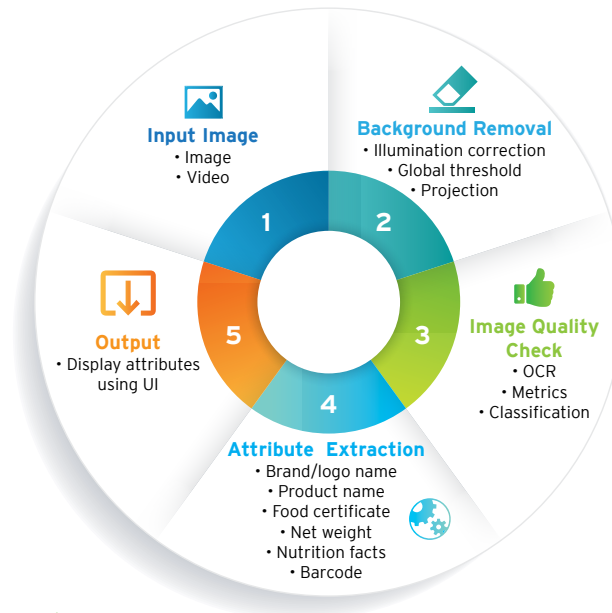
Metadata Extraction: A Five-Step Process

Figure 4

Background Removal

A background removal subsystem removes background color information from different product label images. This is done to improve character recognition accuracy and product label image acceptance. Product label images, with different gradient, solid colors and complex natural scenes, undergo image preprocessing techniques such as background removal to identify and extract the regions of interest. The illumination correction is calculated using morphological operation on gray-scale images.

Edge information is obtained using global threshold techniques. To get the region of interest image, horizontal and vertical projection is applied on the extracted images.

Image Quality Check

The product label images are then put through a document acceptance check subsystem to filter and classify images, which ensures the product attributes are extracted reliably. The subsystem puts the documents into three buckets: "accept," "needs manual intervention" and "reject." Accepted documents can be automatically processed, with no manual intervention.

Because traditional OCR systems don't support cursive/typographic font text, a detection module is used to identify such text regions.

Text region detection is performed by combining maximally stable extremal region (MSER)[4] and Niblack algorithms,[5] which results in low rates of false text detection. Stroke width transform,[6] Euler's number[7] and neighborhood connected-component methods[8] are used to validate character/word regions, and the area of text detection is calculated. Text is extracted using the Nuance OCR engine.[9] The area of OCR character, the mean character height in pixels, and the mean character confidence score are calculated based on text portions occupied by OCR text.

Our proposed system uses the following document acceptance check features as metrics:

- *Detection Percentage* = $\dfrac{\textit{Area of OCR character}}{\textit{Area of text detection}}$

- Mean character height.

- Mean character confidence score.

Classification is carried out by applying a threshold on metrics that result in an automation quality test. Once completed, the process can proceed.

Regions of interest of barcodes, logos/brand names and certification logos are passed to another logo detection and recognition module to detect the brand name and food certifications. Template support is provided to extract different formats of structured text information. Tabular text, volumetric information and other information each pose their own challenges, for which the extraction process is detailed below.

Attribute Extraction

Attributes such as brand name, product title, net weight, food certification, net weight/volume and barcode are extracted from input images with the help of various image processing techniques and the application of AI and NLP.

AI-Enabled Attribute Extraction

Note: Kashi is a registered trademark of Kashi Co.
Figure 5

Brand Name Detection (Logo Detection)

The OCR text output is sent to an NLP engine to receive entity identification. If the logo is not detected, then the process moves to image processing, based on brand name detection. Local and global feature extraction processes are applied to the input image. Features are then transformed to a bag-of-words model[10] using the K-Nearest Neighbors (K-NN) algorithm.[11] The Euclidean distance is calculated from the input image and bag-of-words model to recognize the correct brand name.

Product Title Detection

The product title is found by using NLP-based dictionary management and text similarity.

Standard Certification Detection

Food certification labels such as "USDA" or "gluten-free" are then identified. From the region of interest images, local features are detected. A feature is a part of an image with some special properties that can be used to perform certain calculations, such as tracking and matching. Here, standard food certification matching is applied.

In this example, we rely on local features that describe a part of an image, rather than global features, which describe the image as a whole. Once trained data set features and query image features are found, the process moves to the matching phase. This is performed similarly to how brand name detection is conducted.

Extracting Nutrition Facts

Nutrition Facts:
Serving Size: 2 bars (40g)
Servings Per Container: 6
Calories: 170

Nutrition	Grams	% Daily Values
Total Fat	6	9
Saturated Fat	0.5	3
...	...	...

Ingredients: Whole grain oats, dried cane syrup, rolled whole grain blend, expeller pressed canola oil, soy protein

Distributed by

Note: Kashi is a registered trademark of Kashi Co.
Figure 6

Reading Barcodes

Barcode: UPC -018627030126   100% Recycled   Food Type: Vegetarian

Package: Multi-layer Wrapper
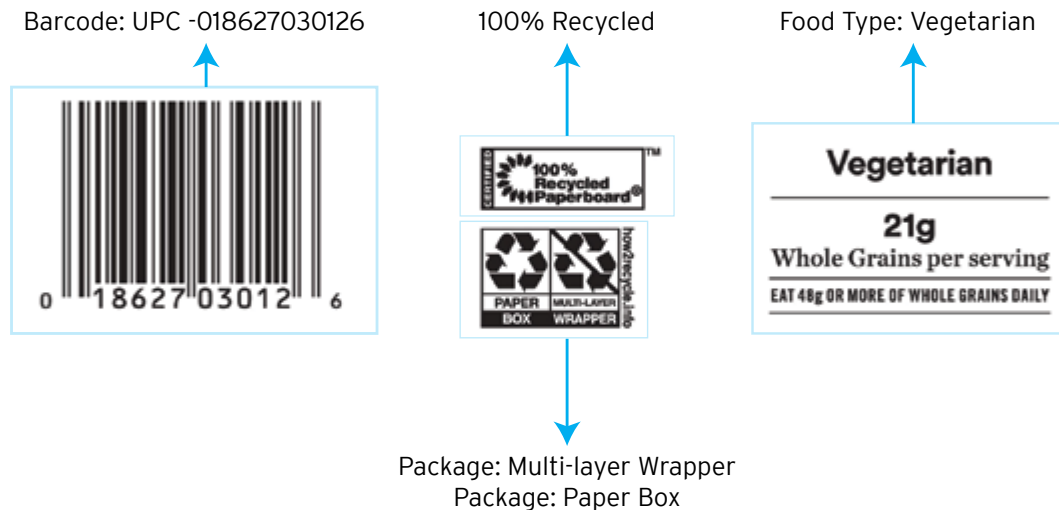Package: Paper Box

Figure 7

Net Weight Detection/Extraction

To extract net weight/volume/quantity, regular expression techniques have been used on text extracted using OCR. Regular expressions have been built using key words related to net weight, quantity and volume.

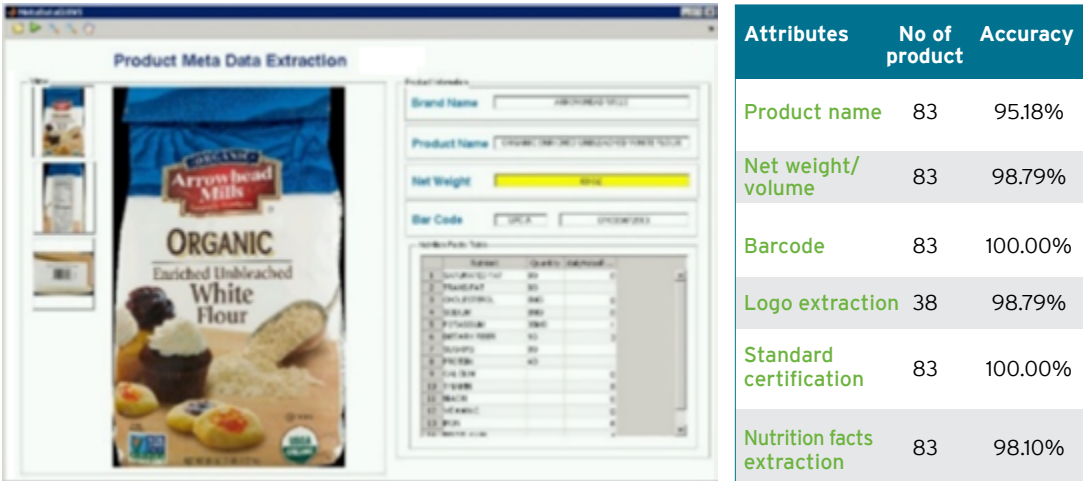### Nutrition Facts Detection and Extraction

As nutrition facts follow a tabular format, morphological operations are used to detect the horizontal and vertical lines. With horizontal lines reference, each text subregion is cropped, and text is extracted using the Nuance OCR. Extracted text is corrected with a predefined vocabulary. A rule-based approach is used on corrected text to extract nutrition data. Figure 6 (previous page) reveals the rule used to extract the nutrition facts from the text.

Barcode Detection

A third-party tool from a standalone library is applied to detect and recognize barcodes. These library functions handle UPC-A format (as revealed in Figure 7).

The Accuracy of Image-Based Product Metadata Extraction

Attributes	No of product	Accuracy
Product name	83	95.18%
Net weight/volume	83	98.79%
Barcode	83	100.00%
Logo extraction	38	98.79%
Standard certification	83	100.00%
Nutrition facts extraction	83	98.10%

Note: Arrowhead Mills is a registered trademark of Hain Celestial Group.

Figure 8

KEY ADVANTAGES OF THIS APPROACH

Product label images are a trusted source of metadata information. Naturally, the process should improve the quality of product metadata and data consistency. Our solution reduces the burden of validating product data provided by various vendors, and provides additional information critical for consumer product discovery, such as brand and certification logos information.

The Results

To assess the performance of our proposed solution, we evaluated it using a real dataset with 352 food products, encompassing 53 brands containing 955 images (including front-, back-, side-view product images). Background removal was used to improve OCR accuracy. We tested with product images with and without background removal, and evaluated character confidence scores and conducted an automation quality check. (See Figure 8 for the results as applied to one particular product.)

THE ROAD AHEAD

Computer vision and AI-based methods show clear potential for applying process automation to reduce data inconsistencies and improve metadata data quality, thereby improving the retail industry's product data capture and metadata extraction processes.

Our proposed computer vision-based approach can be extended to other product categories, such as health, beauty, books, toys and video games. It can also be extended to improve the extraction process for product images with diverse backgrounds, cylindrical and can image labels.

Deep learning techniques can and will be explored to enhance image and text region segmentation accuracy, as well as support cursive and typographic font-based text extraction. Machine learning and NLP techniques can be explored to improve text attribute extraction accuracy.

FOOTNOTES

1   Amy Gesenhues, "Report: E-commerce Accounted for 11.7% of Total Retail Sales in 2016, Up 15.6% Over 2015," Marketing Land, Feb. 20, 2017, https://marketingland.com/report-e-commerce-accounted-11-7-total-retail-sales-2016-15-6-2015-207088.

2   "High-Quality, Complete E-commerce Product Content Increases Sales," BrandView, Sept. 7, 2016, http://www.brandview.com/2016/09/high-quality-complete-ecommerce-product-content-increases-sales/.

3   Andria Cheng, "Consumers Return $642.6 Billion in Goods Each Year," Marketwatch, June 18, 2015, https://www.market-watch.com/story/consumers-return-6426-billion-in-goods-each-year-2015-06-18.

4   Tutorial on MSER, VLFeat.org, http://www.vlfeat.org/overview/mser.html.

5   "Niblack and Sauvola Thresholding," SciKit-Image, http://scikit-image.org/docs/dev/auto_examples/segmentation/plot_niblack_sauvola.html.

6   "Detecting Text in Natural Scenes with Stroke Width Transform," IEEE, http://ieeexplore.ieee.org/document/5540041/.

7   Wikipedia entry on Euler's number: https://en.wikipedia.org/wiki/E_(mathematical_constant).

8   Wikipedia entry on connected-component labeling: https://en.wikipedia.org/wiki/Connected-component_labeling.

9   Nuance website: https://www.nuance.com/print-capture-and-pdf-solutions/pdf-and-document-conversion.html.

10   Jason Brownlee, "A Gentle Introduction to the Bag-of-Words Model," Machine Learning Mastery, Oct. 9, 2017, https://machinelearningmastery.com/gentle-introduction-bag-words-model/.

11   Adi Brohnshtein, "A Quick Introduction to K-Nearest Neighbors Algorithm," Medium, April 11, 2017, https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Gundimeda Venugopal

Lead, Cognitive Computing & Data Sciences Lab

Gundimeda Venugopal leads the Cognitive Computing & Data Sciences Lab's research team within Cognizant's Global Technology Office. He has more than 23 years of IT industry experience in the areas of enterprise architecture, large-scale application and framework development, artificial intelligence, machine learning, NLP, web spidering, information extraction, computer vision, speech processing, biometrics, enterprise search, object-oriented design, web development, middleware, databases/LDAP, performance tuning, embedded systems, networking, protocol design and in-support systems. Venu was the solution architect for multiple large-scale application development projects. He received a B.Tech. in electrical and electronics engineering from J.N.T.U College of Engineering, Kakinada, and an M.Tech. in computer science from Jawaharlal Nehru University, New Delhi. He has filed two patents, written articles for three research publications and won two innovation awards. Venu delivered guest lectures in the areas of digital communications (IIIT, Gwalior) and e-governance (IIM, Bangalore). He can be reached at Venugopal.Gundimeda@cognizant.com | LinkedIn: https://www.linkedin.com/in/venugopalgundimeda/.

Ramakrishnan Viswanathan

Manager, Business Development, Cognitive Computing & Data Sciences Lab

Ramakrishnan Viswanathan is a Manager of Business Development in Cognizant's Cognitive Computing & Data Sciences Lab within the company's Global Technology Office, focusing on AI ML services, cognitive technology and emerging technologies. In addition to his time spent in GTO and Cognizant's Application Value Management Practice, he has over 13 years of experience in pre-sales, strategic partnerships, business development, client relations and business analysis. Ram has an executive program in sales and marketing (EPSM) degree from Indian Institute of Management, Calcutta (IIMC). He can be reached at Ramakrishnan.Viswanathan3@cognizant.com | LinkedIn: www.linkedin.com/in/ramakrishnanviswanathan.

Rajkumar Joseph

Architect, Cognitive Computing & Data Sciences Lab

Rajkumar Joseph is an Architect within Cognizant's Cognitive Computing & Data Sciences Lab within the company's Global Technology Office. He has over nine years of experience in innovation, research and product development in the field of computer vision, artificial intelligence, data science, mobile computing and IoT. He received an M.Tech in industrial mathematics and scientific computing from Indian Institute of Technology Madras (IITM) and completed an executive program in business analytics (EPBA) from Indian Institute of Management, Calcutta (IIMC). He can be reached at Rajkumar.Joseph@cognizant.com | LinkedIn: https://www.linkedin.com/in/rajkumar-j/.

Naresh Babu N T

Technology Specialist, Cognitive Computing & Data Sciences Lab

Naresh Babu N T is a Technology Specialist within Cognizant's Cognitive Computing & Data Sciences Lab within Cognizant's Global Technology Office. He has eight years of experience in research and software development, embedded platform and data analysis. In addition to image and signal processing, his areas of interest include computer vision, machine learning, pattern recognition and soft computing. Naresh received an M. S (By Research) from Anna University (MIT Campus), where he specialized in signal and image processing. He can be reached at Nareshbabu.nt@cognizant.com | LinkedIn: https://www.linkedin.com/in/naresh-babu-n-t-79086419/.

Cognizant

COGNIZANT'S COGNITIVE COMPUTING & DATA SCIENCES LAB

The Global Technology Office (GTO) is the core technology organization of Cognizant, with a mission to power and accelerate our capabilities to harness transformative technologies that enable our people, customers and processes to navigate the shift in the work ahead. As part of GTO, the Cognitive Computing & Data Sciences (CDS) Lab's vision is to explore emerging and cognitive technology areas in artificial intelligence, machine learning, natural language processing, voice/speech recognition and computer vision. The CDS lab builds innovative industry-specific cognitive platforms and solutions for digital business transformation.

ABOUT COGNIZANT

Cognizant (NASDAQ-100: CTSH) is one of the world's leading professional services companies, transforming clients' business, operating and technology models for the digital era. Our unique industry-based, consultative approach helps clients envision, build and run more innovative and efficient businesses. Headquartered in the U.S., Cognizant is ranked 205 on the Fortune 500 and is consistently listed among the most admired companies in the world. Learn how Cognizant helps clients lead with digital at www.cognizant.com or follow us @Cognizant.

World Headquarters

500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277

European Headquarters

1 Kingdom Street
Paddington Central
London W2 6BD England
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102

India Operations Headquarters

#5/535 Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060

TL Codex 3263