Dissertation Title:

New Horizons: Exploring a Decade of London New Build Development with Convolutional Neural Networks

Name:
Anthony Sutton

Programme:
MRes Spatial Data Science & Visualization

Institution:
UCL

Date:
1st September, 2018

Word Count: 8414

Version 1.1	01/09/2018	WC: 8414	Final Submission
Version 1.2	01/04/2019	WC: 14803	Workflow Sections added
Version 1.3	15/07/2019	WC: 20988	Extended Study Sections added (Grad Cam, Keras Fine Tuning)

Declaration

I, Anthony Sutton, herby declare that this dissertation is all my own original work and that all sources have been acknowledged.

“..the meaning of a word is its use in the language”

Philosophical Investigations -   Wittgenstein

Dedication:

Dr Sarah Wise – DS Tutor, Mentor and Support
London Borough of Tower Hamlets, Planning Department, Data and GIS Team – Sponsorship, Expertise & use of ESRI SDE platform

Abstract

Fuzzy Labeling controls are applied to London Property and New Build Address Datasets for defining practical Machine Learning friendly class types for a Convolutional Neural Networks Image Classification task on London Street View Building Façade Imagery Data.

Using a generalized workflow proposed by Kang et al and originally applied to US and Canadian cities, we found that the London Dataset was harder to achieve as clear cut classification results upon, possibly reflecting a greater element of high density and mixed use tenure as an increasingly defining geo-informational characteristic of the London urban terrain.

Several ubiquitous workflow software tools were built to establish a high level of confidence in the data accuracy, geo referencing and model validation components of the study. Particular attention is given to a comprehensive application and exploration of current CNN Classification methods.

This approach allows our study to make several observations on the nature of the urban environment being modeled, the challenges faced in applying AI and ML to the urban environment and also to provide discursive material for both the computational and representational camps that delineate prominently in today's ongoing cognitive debate.

Table of Contents

List of Figures and Tables

1 Research Goal

To identify and examine the Typological character of London New Build Development through the application of current Convolutional Neural Network Image Classification Techniques.

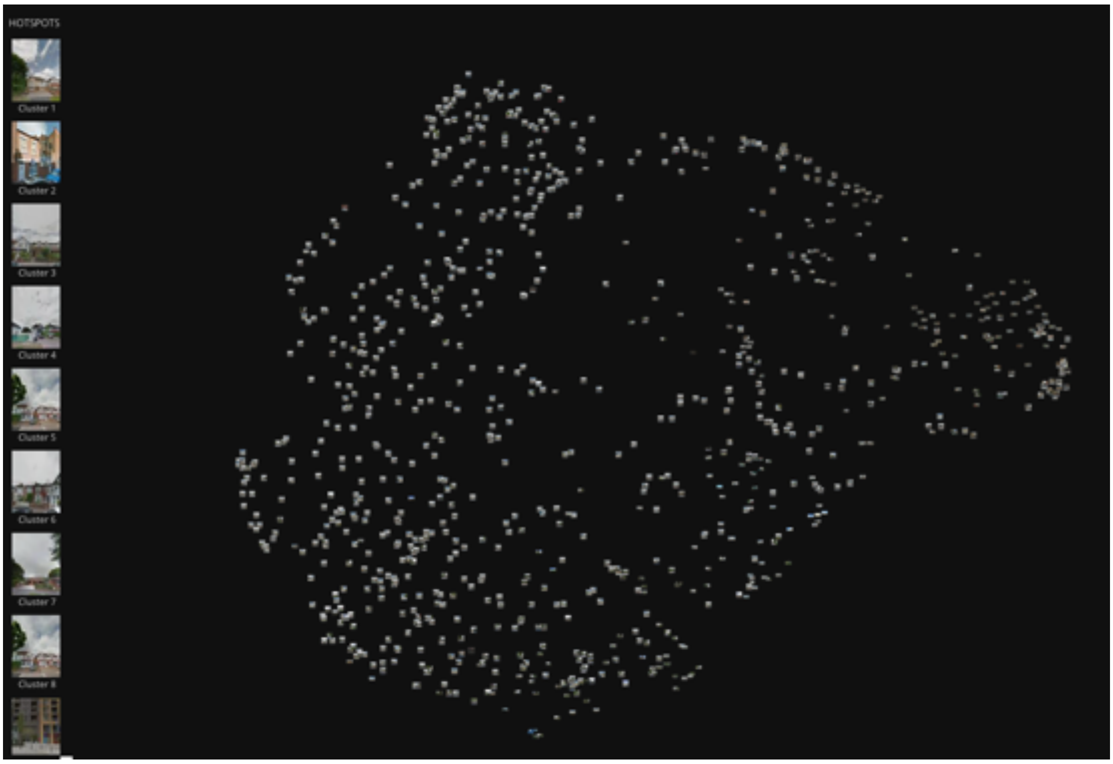


Fig 1.1 *– Inception v3 CNN Image Vectors for London new build Flat, House, Industrial and Office Building Types analyzed with a TSNE Manifold learning data embedding technique. A Flats cluster is located to the far right.*

Aim

This study's motivation is to enquire whether current machine learning and computer vision techniques can provide insight into the nature of London's new build development.

Our data model is comprised of the London Development Database, the Google Street view dataset and the Ordnance Survey Address Base dataset. The completed model provides us with completion dates, development description, analytics on density (unit, area) linked to a street level digital image of the property and joined with an OS 3 tier property type classification.

London Development Database (LDD) and The London Plan

New development that involves new units and floor space is recorded in the LDD, which is used as an analytical tool in the London Plan making and implementation process. It contains data dating back to about 2004. With completions taking time to be recorded by Local Authorities and (on larger developments) in reaching phased completion, the current data is strongest from about 2006 and 2016.

Two important themes in the challenge and purpose of the London Plan Process are those of Density and Consent. If increasing *density* and increasingly limited resources to accommodate this change is a primary concern in the London Plan (Katz et al – 2016), then one might argue that *consent* in how it deals with this challenge and in the process that it embodies is of an equally important nature. Machine Learning provides us with a set of tools that are often primarily concerned with the optimization of loss and cost.

Whilst the complexities of the UK Planning process are outside the scope of this study we note that technical innovation that offers up the possibility of a more empirically aggregate audit, analysis and assimilation of the cadaster might ultimately offer new means for the optimization/reconciliation of agent and agenda enfranchisement that, for better or for worse, struggle and compete on an increasingly ever growing and expansive scale.

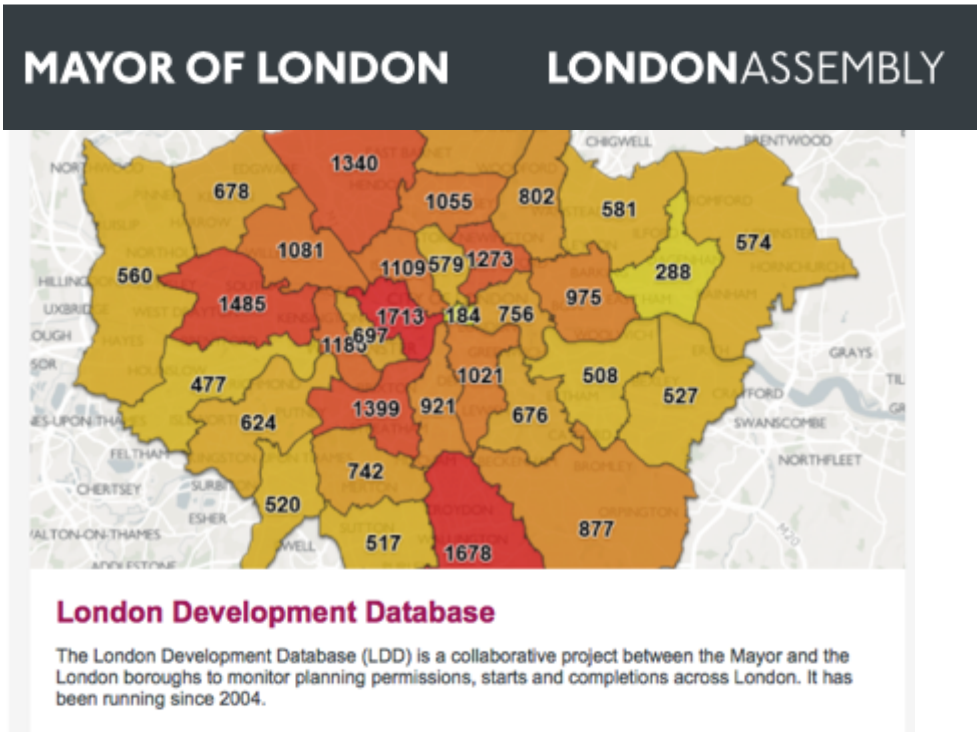


Fig 1.2 *– The Greater London Authorities(GLA) London Development Database(LDD)*

Method

The overall method comprised of two parts. Firstly, to apply a Convolutional Neural Network based Classification workflow on the dominant property types found in the LDD. Secondly to use the feature description element of the above method to examine further the individual and aggregate nature of these types.

The AddressBase Premium dataset which contains property classifications is a licensed dataset. We acquired limited access through our employers, who helped in joining the dataset for the study using ArcMap/ESRI SDE, via a spatial buffer join, on LDD BNG coordinates, returning the top matched address. The Georeferenced data was projected to WGS 84 Web Mercator for use with the Google street view API. As of August 2018, the latter is also now a licensed dataset. Our download was completed just prior to this and could have benefited from more time to experiment with the radius and panoramic parameters available through the API to provide more dynamic control of the label selection of candidate images.

2 Literature Review

2.1 What does New London look like?

How can we best see and assimilate the scale of urban growth presently encountered by our major urban megalopolis? Can Artificial Intelligence assist us in this challenge and help us to see and analyze the visual nature of this growth and change? What are we looking at and what does it mean when we attempt to label, categorize and describe the elements of our urban spaces that exist in a continuum of production?

2.2 AI in Urban Design and Planning

Wu and Silva outline a comprehensive overview of AI applications in the field of planning and urban land dynamics. They group these solutions into 4 work streams focusing on Artificial Life, Intelligent Stochastic Optimization processes, Evolution computing and Spatial DNA, and Knowledge based systems (Wu et al – 2100).

The emergent idea of their review is an emphasis on the potential of hybrid applications of the above mentioned four disciplines.

We see this idea echoed in the adoption of Ensemble methods in the domain of Machine Learning as an increasingly popular and effective way of achieving greater precision and results to the tasks at hand (Geron - 2017).

Batty outlines how and why the use of science can be advanced in planning describing how it has retreated from theories about how cities function in favor of deriving models that offer conditional predictions.

2.3 The Role of Vision in Cognitive Science

From Informational Processing to Anti Representational Models of the Mind

Vision has played an important role in Cognitive Science as an intermediary link between perception and process and in examining this relationship has brought forth research that is interested in the human and the machine. Vision has played a frequent and important role in several key experiments and ensuing theories that have arisen from the cognitive scientific approach.

An empirically rich relationship, it provides the motivational basis for the Information Processing modeling approach to understanding Cognition(Bermudez – 2014) and provides insight to and from the phenomenological and epistemological philosophical domains, whilst focusing on building models of the mind and intelligence, be it artificial or otherwise.

"A physical symbol system has the necessary and sufficient means for general intelligent action."

— Allen Newell , Herbert A. Simon

David Marr's approach to Vision was based on the assumption that the Vision process infers 3D representation based on the information provided by the eyes. Marr decomposed the Vision task into hierarchical levels. The functional level where the task can be understood geometrically, the algorithmic level and the hardware or neuronal level. Vision and Vision Experiments treated in this way propose computational models of the mind as a promising and valid means of understanding mental processes and experience.

From its origins through to the biologically inspired information processing models of neural networks and in particular convolutional neural networks which we expand in detail later, Vision has also gone on to play an important role in more recent developments in Cognitive Science such as the ongoing delineation between Representational vs. Elimitivist Models and theories in the philosophy of mind.

Gibson rejected the conception of Vision as an inversion of the imaging process and refuted the representational approach to Vision. Gibson assumed that what is perceived, the phenomenology of Vision, can be explained by the dynamic optical array. The images on the retinas of the eye sample the dynamic optical array at locations of different orientation. Gibson drew a connection between the needs of the organism to aspects of the stimulus directly. In Gibsonian Ecological psychology “places perception and cognition” are to be understandable only in terms of action in the environment.

Situated cognition is a theory that posits that knowing is inseparable from doing by arguing that all knowledge is situated in activity bound to social, cultural and physical contexts(Greeno & Moore)

Recent investigations of learning challenge this separation of what is learned from how it is learned and how it is used. The activity in which knowledge is developed and deployed, it is now argued, is not separable from or ancillary to learning and cognition. Nor is it neutral. Rather, it is an integral part of what is learned. Situations might be said to co-produce knowledge through activity. Learning and cognition, it is now possible to argue, are fundamentally situated(Brown et al).

“Dynamical models of Cognition like Sussex Robots show that significant portions of cognition might be explained without mental gymnastics”

— Chemero, Radical Embodied Cogntive Science

Chemero argues that cognition should be described in terms of agent-environment dynamics rather than in terms of computation and representation.

Complex dynamical systems exhibit three key characteristics (Gallagher and Appenzeller, 1999). First, they consist of a number of interacting components or agents. These components can be homogeneous or heterogeneous. A collection of cortical areas or simple artificial agents can comprise a homogeneous, complex dynamical system; a brain in a body in an environment can comprise a heterogeneous, complex dynamical system. A second property is that these systems exhibit emergent behavior in that their collective behavior exhibits a coherent pattern that could not be predicted from the behavior of the components separately. Third, and most importantly, this emergent behavior is self-organized in that it does not result from a controlling component agent. These three characteristics can be seen clearly in phenomena such as bird flocking.

Contemporary methods of non-linear dynamics embrace the complexity of self-organized behavior and, accordingly, can provide deep insights about the behavior of real-world time-evolving processes.

A non-linear system is one in which the system’s output is not directly proportional to the input, as opposed to a linear system in which the output can be simply represented as a weighted sum of input components. Complex dynamical systems are non-linear in this sense, so their behavior is never merely the sum of the behavior of the components (Van Orden, Holden, and Turvey, 2003).

Event-related potentials (ERPs) are very small voltages generated in the brain structures in response to specific events or stimuli (Blackwood and Muir, 1990). They are EEG changes that are time locked to sensory, motor or cognitive events that provide a safe and noninvasive approach to study psychophysiological correlates of mental processes. Event-related potentials can be elicited by a wide variety of sensory, cognitive or motor events(Shravani Sur and V. K. Sinha1 2009).

Spiking neural networks SNNs, are often referred to as the 3rd generation of neural networks. They operate using spikes, which are *discrete events that take place at points in time, rather than as continuous values*. Highly inspired from natural computing in the brain and recent advances in neurosciences, they derive their strength and interest from an accurate modeling of synaptic interactions between neurons, taking into account the time of spike firing. SNNs overcome the computational power of neural networks made of threshold or sigmoidal units. Based on dynamic event-driven processing, they open up new horizons for developing models with an exponential capacity of memorizing and a strong ability to fast adaptation (Moisy- Bohte 2012).

This study has granted us experience in applying AI to a real world challenge, specifically symbolic neural network based AI. In the study's exploration of building typology it has also allowed us to observe the role of semantics, both visual and conceptual and its impact on using connectivist tools to make sense of a human and urban environment.

Our study's interest in Modern Building Typologies and the application of a Convolutional Neural Network to a real world Image Classification question grounds the work firmly in the Symbolic AI domain, however our aim is to also allow its outcomes and findings to be read in a context that informs an opinion on Chemero's proposal of a need for a "Radical Embodied Cognitive Science". This will stem from our research's dealings with visual typology and with the lessons learned approach to applying an ML analysis workflow in the exploration of urban scientific domain specific questions. In order to do so we will begin with going over some of the basics of Computer Vision.

2.4 A brief recap of the Computer Vision Approach from SIFT to CNN

Computer Vision techniques that process and analyze images typically involve digitizing an image into numerical data form and extracting features of interest into vector space in order to smooth noise and to find edges and other image patterns(Forsyth - 2003).

Machine Learning methods can then be applied to the visual subject's reduced feature space to approach image classification goals such as pattern recognition or object detection.

Simple modifications of histogram Feature descriptors such as that applied in the Histogram of Gradient orientations(HOG) technique reduce the image to orientation rather than edges to achieve greater contrast between an image subject and its background. An earlier version of the HOG feature using a Scale Invariant Feature Transformation (SIFT) uses feature vectors which are less prone to distortion when translated and that recognize objects by matching and verifying points in the image to a pose estimation model(Norvig -2016).

A great deal of variation and diversification has developed on these basic principles in the field of Computer Vision.

More recently interesting variations in the basic architecture of Deep learning algorithms and Convolutional Neural Networks have allowed for some impressive Classification Challenge results and resonated with the growing public interest in understanding Artificial Intelligence(AI), in particular Neural Networks.

Neural Networks (described in detail later)whilst initially generating a lot of optimism among proponents of AI was found to reach an impasse in development notably as regards the XOR Problem(Beale –1990). The application of backwards propagation allowed for this impasse to be broken and for research to proceed.

Most recently they have been achieving increasingly impressive and remarkable recall accuracy rats. The prediction error rate error rate in the ongoing movement of competitions(e.g. the ILSVRC Image net challenge, for example) has been improving to over 90% which has beaten Human Levels of performance in the same field.

2.5 Computer Vision based Research into the Urban Environment

Research that applies computer vision and machine learning techniques to the urban environment will be concerned with detecting change in urban physical forms and morphology or with an analysis of the properties or nature of the artifacts represented in the image. They might utilize either remote sensory images or ground level based imagery.

Change detection methods can be grouped into change enhancement techniques and nature of change techniques(Singh - 1989). Enhancement techniques typically locate "changes and the magnitude of changes but do not deliver the information of the nature of change"(Chan et al - 2001). For the latter a further classification process is needed.

In 2001 (Chan et al) applied tested four classification algorithms(Multi Layer Perceptron, Learning Vector Quantization and Decision Tree and Maximum Likelihood Classifiers)comparing recognition rates, ease of use on simulated training data, hard to obtain at the time. Since then advances have been made in the availability of remote sensed data, and whilst LVW came first in their rankings, MLP was ranked second and they highlighted the cost and difficulty posed in setting up the Neural Network.

Another response to the then lack of training remote sensed image data was taken by (Oh - 2006) who applied an Unsupervised Classification or active learning approach into identifying a Main Street, urban setting, based on their spatial and sematic relationships over building geometry. Here we see AI being used to develop data analysis and inference capabilities whose unprecedented value lies in its scale being large.

Liu et al harnessed the abundance of now freely available (in their case from China's Baidu map api) in their hand labeled evaluation large scale change to the urban visual environment using both SIFT histograms and deep convolutional networks. For the latter, a SVM classifier was trained on a pre trained neural network (AlexNet and GoogleNet). They used a body of architectural expertise in developing an expert rating scheme relating to Visual Quality and Visual Continuity and highlights the qualitative and quantitative divide faced by Computer Vision research in a Social Urban context(Lio - 2017).

Naik et al approached the perceptual implications of vision based urban research by using crowd sourced labeled data(from the MIT Place Pulse project) as the measure of change by which to correlate and predict characteristics of neighborhood improvement(Naik et al – 2017) . Their streetscore system allowed them to build a high resolution dataset of physical urban change for 5 major US cities, and found that in agreement with economic theories of human capital agglomeration, neighborhoods that are densely populated by highly educated individuals are more likely to experience positive urban change and link this to Invasion theory and neighborhood tipping theory.

(Porzi et al – 2015) had developed this approach with the same Place Pulse labeled data by formulating the problem of quantifying human perception as a ranking problem. They use this ranking with a CNN trained from scratch relying on a pooling layer fed to by latent detectors which aimed to discover small image regions.

A more quantitative and Unsupervised object discovery approach to the problem of defining the visual nature of an urban locale, was taken by (Doersch et al - 2015) in their paper What Makes Paris look like Paris? Building on work by Shirivistra et al in adapting the distance metric using discriminative learning a linear SVM detector is applied for each visual element(extracted as patches of images from the Google Street view depository). The procedure produces a weight vector which corresponds to a new, per element similarity measure which aims to be both repetitive and geographically discriminative. This results in extracting geo informative elements that are arguably able to represent a stylistic representation of city.

Along similar lines research has been made into classification of Building Instances and Architectural styles. (Kang et al – 2018)propose a workflow based on CNNs which classify façade images from Google street view and remote sensing images. The method was applied to generating building classification maps on both region and city scales of several cities in Canada and the US. Their method employed a Transfer Learning technique, whereby the penultimate layer of a pre trained CNN network is extracted and trained to a subsequent new image training set.

Less generically (Shalnuts et al – 2011) venture into a classification of façade windows according to architectural style based on clustering and learning of local features. Rather than using different texture features or shape descriptors, they incorporate texture and gradients into their use of image descriptor using a Scale Invariant Feature Transform to gather information of gradient directions which they use to use unsupervised kmeans clustering to learn a visual vocabulary(codebook).

(Feng et al – 2015) apply the bag of visual words model to deep spatial features for the complexity and diversity specifics of a geographical scene classification, and has particular utility to where landscape and land cover patterns have an urban setting or mirroring.

Another approach to automatic architectural style recognition is employed by (Mathias et al. – 2011) Differing from the tendency for many algorithmic approaches to take an inverse procedural modeling direction it approaches the image of the façade as both a scene classification and image rectification problem. They highlight a practical issue commonly faced when dealing with Street View data and base a solution on a façade splitting approach.

Research into architectural façades as aesthetic objects might also find affinity in approach with several computer vision methods that have been applied to fine art painting classification. Arora and Elgammal's comparative study(2012) reviews the performance of discriminative vs. generative models. They also compare features level research into semantic level vs. low level and intermediate features present in the painting.

Focusing on a intrinsically Large Classification task (Saleh et al – 2015) investigate a wide ranging list of visual features and metric learning approaches to learn an optimized similarity measure.

Building on the fundamental computer vision idea of feature extraction, (Nevedova and Larek - 2017) investigate the 2 dimensional domain of the characteristics of temperature and harmony, with a view to extending the algorithmical products of this approach into generative realms pointing towards new potential areas of research for modeling the cognitive creative process in the AL and ML domains.

The Google art project uses TSNE dimensionality reduction to similar effect and presents us with a single aggregate interface from which to visualize and access and explore assimilation routes into the full body of content or canon of artistic endeavor available in a digital format.

Our Investigation into the visual properties of an object, as outlined hitherto, are concerned with form or shape or scene analysis but research can also stem from a micro level approach, for example analysis of an object's constituent textural properties. (Xiao et al – 2010)present a method to extract feature by Principal Component Analysis. Candidate features of interest are highlighted by applying different luminance directions to capture the particular view of a given surface from a fixed observation point with the subsequent series of images captured and analyzed. Signals of significance are gained by using variable illumination. This approach it transpires can be used to automatically detect and classify texture feature defects.

The division of macro or macro level approaches to visual analysis highlights the role of boundary definition of feature space scope in determining the type of research outcome. This touches on the cognitive scientific question that seeks clarity in the definition of what it means for an object to exist visually in the mind.

Several important visual databases designed for use in visual object research have been central to the development of the field computer vision and should also be mentioned in terms of how they might relate to Urban based research.

In 2010 The goal of the SUN database project is to provide with a comprehensive collection of annotated images covering a large variety of environmental scenes, places and the objects within(Xiao -2010).

MIT's Place365 can be used to train artificial systems for high-level visual understanding tasks, such as scene context, scene recognition (Zhou 2017)

Coco (Common Objects in Context) presents a new dataset with the goal of advancing the state-of-the-art in object recognition by placing the question of object recognition in the context of the broader question of scene understanding.

Cityscapes(Cordts et al - 2016): For semantic urban scene understanding, however, no current dataset adequately captures the complexity of real-world urban scenes. To address this, we introduce Cityscapes, a benchmark suite and large-scale dataset to train and test approaches for pixel-level and instance-level semantic labeling.

Mapillary(Neuhold – 2017) is the world's largest and most diverse publicly available, pixel-accurately and instance-specifically annotated street-level imagery dataset for empowering autonomous mobility and transport at the global scale.

Another use of CNNs is DELF (Deep Local Feature). This based on convolutional neural networks, but which are trained on image-level annotations on a landmark image dataset. A large-scale dataset, referred to as Google-Landmarks dataset, is used which involves challenges in both database and query such as background clutter, partial occlusion, multiple landmarks, objects in variable scales(Noh -2017)

Closely related is the task of semantic segmentation, a task which involves segmenting an image into all its found objects. This is similar to object detection, but actually demarcates the full border of each found object, rather than just its bounding box.

Semantic segmentation and object detection have only become feasible relatively recently. One of the major limitations holding them back previously, besides the increased complexity compared to single-class classification, was a lack of available data. Even the larger scale ImageNet dataset was unable to do anything about detection or segmentation because it had sparse information about the locations of objects. The more recent datasets like MS-COCO have added richer information for each image into their schema, enabling localization, detection, and segmentation to be pursued in greater detail.

2.6 Typology

Having looked at the tools and means at our disposal for collecting, representing and processing Urban data both in terms of large scale and granular detail we now turn to the issue of how we might classify, interpret and share this information.

Typology is a key concept in describing Urban Form, Architecture and Society.

Moneo describes it as a concept that describes a group of objects characterized by the same formal structure and is fundamentally based on the possibility of grouping objects by certain inherent structural similarities. The relevance of this approach from the cognitive and AI perspective is that architectural objects are not only described by types but produced through them(Moneo - 1978).

The Italian architect Aldo Rossi mirrors this concern with work ebbs and flows and focused on the synthesis of a variety of differential aspect as an expression for the search for prototypical form(Aldo Rossi – 1966).

We see related ideas regarding the flexible integration of fractured elements into a common continuing transforming but singular whole reflected through out the field.

Venturi reflects this conceptualization of type as something transparent, by focusing on the image as the type. By seeing architecture as Symbol and Space, the urban types take on a linguistic nature, and the urban form becomes a living medium through which we can see archetypal patterns of communication, and the typologies they belie(Venturi et al - 1966).

The notion of looking into and beyond the artifice is reflected by Lynch in What Time is This Place? where urban form is described in terms of “Change Made Visible”. Importantly these are not trivial or random changes but changes that reflect archetypal elements and flows to the urban population, be it from circadian body clocks or in the urban expression of our past, present and future.

Batty expresses these flows in more formal terms, describing them as summations of what happens at locations. This allows him to model these flows e.g. via gravitational models and spatial interaction theory. “Departure of symmetry and a symmetry of the reality” (Batty - 2013)

(Castells – 2000) expresses a similar idea in The Rise of the Network Society. “Space is the expression of society” and the new logic underlying such new spatial forms and processes. Structural Transformation. Space is the material support of time sharing social practices. The change is so flexible he proposes the idea of Timeless Time and the changing relationship between time and society esp. under the auspices of the new information paradigm facing society and culture.

This idea of timeless time is also reflected in Alexander's The Timeless Way of Building(Alexander - 1979). Bring order out of nothing but ourselves, happens of its own accord. He applies this motivation to good and bad design. It is this idea of components to the whole, that are allowed t. The single quality which makes the difference cannot be named. A qualitative approach to Unity and togetherness be it in design, architecture or place building.

In a sense this reflects Wittgenstein's "Familial Resemblance". The meaning of a word is the sum total of its possible uses:

"but there need not be one thing that is associated exclusively with a particular word and is at the same time common to all its possible uses"(Magee -1987).

He differentiates between meaning and the use:

Using a refutation of the Augustinian theory of language (the individual words in language name objects - sentences are a combination of such names) the thesis that meaning-is-use is the best overall explanation of all the relevant meaning-facts or meaning-phenomena. From this it follows that the meaning-is-use thesis is true, including the important qualification that sometimes the human act of ostending an object that bears a name also explains the meaning of that name (Bearsley-1983)

So we see Activity as central to Wittgenstein's account of meaning, whether or not that account is a theory, description, therapy or, simply, an investigation

One could argue that Chomsky's theory of Universal Grammar(UG) resonates with this key tenet in its approach to the genetic component of the language faculty. The basic postulate of UG is that a certain set of structural rules are innate to humans, independent of sensory experience.

Chomsky(1956) argued for " A Language to be a set(finite or infinite) of sentences, each in finite length and constituted of a finite set of elements".

Generative grammar refers to an explicit, formal characterization of the (largely implicit) knowledge determining the formal aspect of all kinds of language behavior. The basic aim of transformational grammar is to explain the principles that state the acceptable ways of transforming deep structure of linguistic elements (Bierwisch -2001).

The 'Generative Enterprise,' as the program was called in Chomsky (1982), went through a surprising development with remarkable changes and modifications, pursuing however the completely coherent and constant research strategy, attempting to characterize and to explain the properties of linguistic knowledge. Each stage of this research program extended, deepened, and simplified previous results and insights and was accompanied by consequences within the narrower domain of linguistics as well as the wider range of related fields, such as ontogenetic development, psychology of language use, problems of biological evolution, or philosophy of mind (Bierwisch -2001).

Both Wittgenstein's and Chomsky's work relate fundamentally to our cognitive motivation of understanding the urban environment and to the notion of visual semantics which feature throughout the study of Computer Vision.

Returning to the field of Urban Studies, Lefebvre (1970) describes this formal and rhythmic flow of meaning distinctly in urban terms in "The Urban Revolution". Attempting to illustrate a depth of crisis in the uncertainty and perplexity of this "Critical phase" entrapped by the blind fields of the newly emergent present the unprecedented realization in the blind fields of everyday living. Topological properties of urban space, properties or system of pertinent oppositions: the open and the closed, dominated and the residual, symmetric and asymmetric; public and the private.

Dovey and Woods (2011) reflect this idea in urban design terms, presenting a typology for the mapping and analysis of urban interfaces, public and private. A Socio-spatial assemblage wherein types are diagrams of connectivity that enable the creation, production and reproduction of ideas, goods, services and identities.

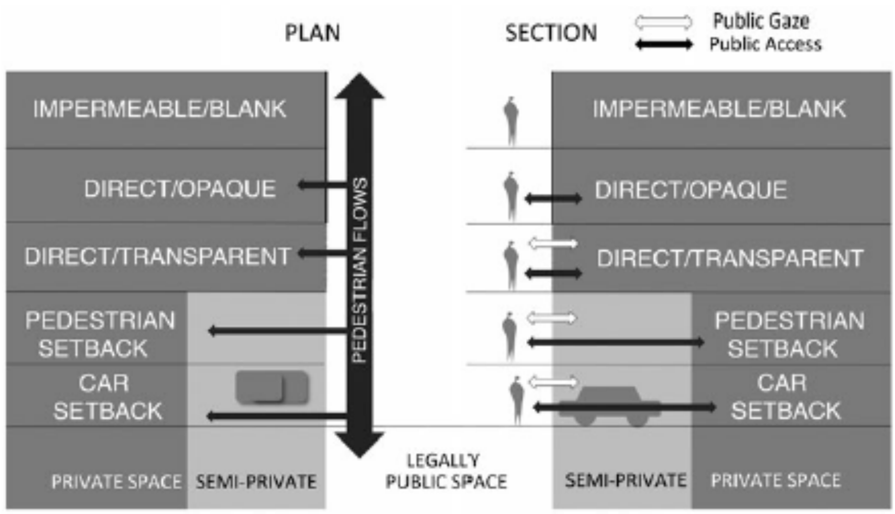


***Fig 2.1**– Dovey and Woods – Urban Interface Typology*

Their typology provides an effective formal framework from which to examine the nature of the spaces that are in production in urban design.
The also see this typology in terms of combinations of setback, transparency, access and mode of access, and contextualize their study in the territory of complex adaptive systems that incorporate ideas of emergence and self organization. However they are not overwhelmed with connation of predictability and systematic control, but rather see complex adaptive

assemblage in the analysis of the mixed morphology of the Australian inner city, responding to multiple factors including urban design codes. The interface was famously interpreted by Jacobs in terms of eyes on the streets and the uses of sidewalks(Jacobs - 1961).

Typology therefore allows for a methodic study of a dynamic process that 'constantly undergoes modifications not only in terms of producing(and reproducing) social relations, but also continually influencing spatial adaptions in newly built forms"(Law et al 2007).

Law highlights the different ideas held by residents about what community means and the difficulties and ambiguities implementing a community driven policy agenda. Rogan extends this study of the provision of public and private space into the increasing effects of Density on the use of housing typology using studies of Low, High and Super High Density development trends as case studies. Highlighting the emergence of the differential attributes of the typologies of the last decade and how they are effected by the London plan process and the focus on high density, mixed tenure housing, sustainable communities and social inclusion.

In "Buildings and Power" (Markus – 1993) puts the emphasis on buildings as not primarily art, technical or investment objects but as social objects, placing the meaning and understanding of buildings in the context of power relations or flows. He distinguishes between two kinds of relations. First, those of power, based on the distribution of limited resources and Second, those of bonds which permeate and subvert all social relations. This analysis of historical text, context the subjective experience of buildings themselves presents the building as a shaper of these power relations between and amongst the social groups and processes they conduct, through their application of societal concepts in form, function and space.

To Do: Synthesis of Foucault's Artifacts from The Archaeology of Knowledge with study's enquiry into Visual Grammar of the Built Environment.

2.7 The Image of the City and Blind Fields

In Image of the City Lynch argues that for any given city, a corresponding set of mental images exist in the minds of the people who experience that city. Contributing to those images are five qualities which Lynch identifies as Paths, Edges, Districts, Nodes, and Landmarks.

in 1970, the French sociologist and philosopher Henri Lefebvre described the "blind field" as a particular experience of urban existence, in which the culturally or politically constructed concept of a place obscures the actual experience of daily life.

Just as the human eye's blind spot is subjective, the urban blind field too is dynamic and shifting. Looking from multiple viewpoints is necessary to the blind field's exposure and more genuine portrayal(Beltrano)

2.8 From Types to Design

Alexander uses theories of type, and principles from set theory to underpin his approach to the addressing fundamental challenges to achieving sustainable and harmonious design. In Notes on the Synthesis of form (Alexander – 1964) he frames good and bad design solutions in terms of good fit and misfit elements within a greater inter connected whole. Self Conscious or misfit design solutions tend to lack fitness between the form of the design solution and its context. He goes onto develop this theory in A Pattern Language and The Timeless Way of Building, experimenting with a proposed fundamentally interconnected approach to design and its elements based on definitive behaviors of geometric patterns in space.

2.9 Design as Consent

In a “Theory of Markovian design machines” Batty (1974) examines in depth the idea of hierarchy used as a sequence in which to synthesize or resolve competing factors toward an optimal solution. He underlines how Alexander’s understanding of the process of synthesis is not just a process of averaging, but one in which the emerging solution anticipates the form of structure. He goes on to apply the Markov process from probability theory to formalize optimal selection in design problem resolution. A process satisfies the Markov property if one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history, hence independently from such history, that is, conditional on the present state of the system, its future and past states are independent. (Rozanov – 2012)

In Machine Learning, Reinforcement learning is where an agent to optimize rewards based on the actions taken from observations made within an environment. Markov decision processes(Based on the Stochastic processes with no memory called Markov chains) feature in Reinforcement Learning(Ageron-2017) and are increasingly easier to implement in Machine Learning platform libraries.

Batty applies Markov process to worked examples of Local Authority resource allocation(Batty -2013) between vested interests and agents, and thus it is here that we see a formal generalized approach to the topic posited earlier in our Study’s motivational theme of Decision optimization, or the optimized convergence of consent.

Figure 4. The Alexander–Manheim hierarchical design solution.

Fig 2.2 *– Batty's Markovian Design machine (below) applied to the Alexendar Manheim Highway location Problem(above)*

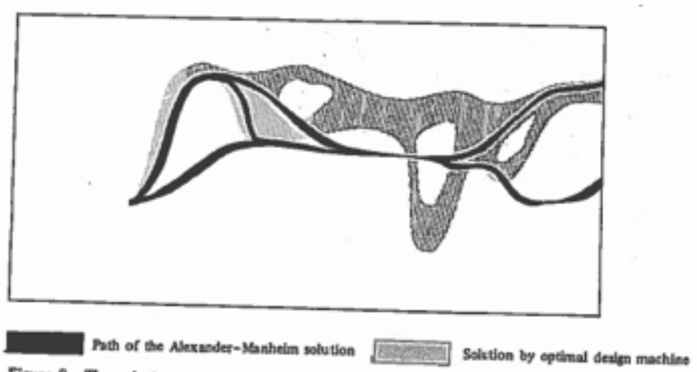


Figure 8. The solution generated by the optimal design machine.

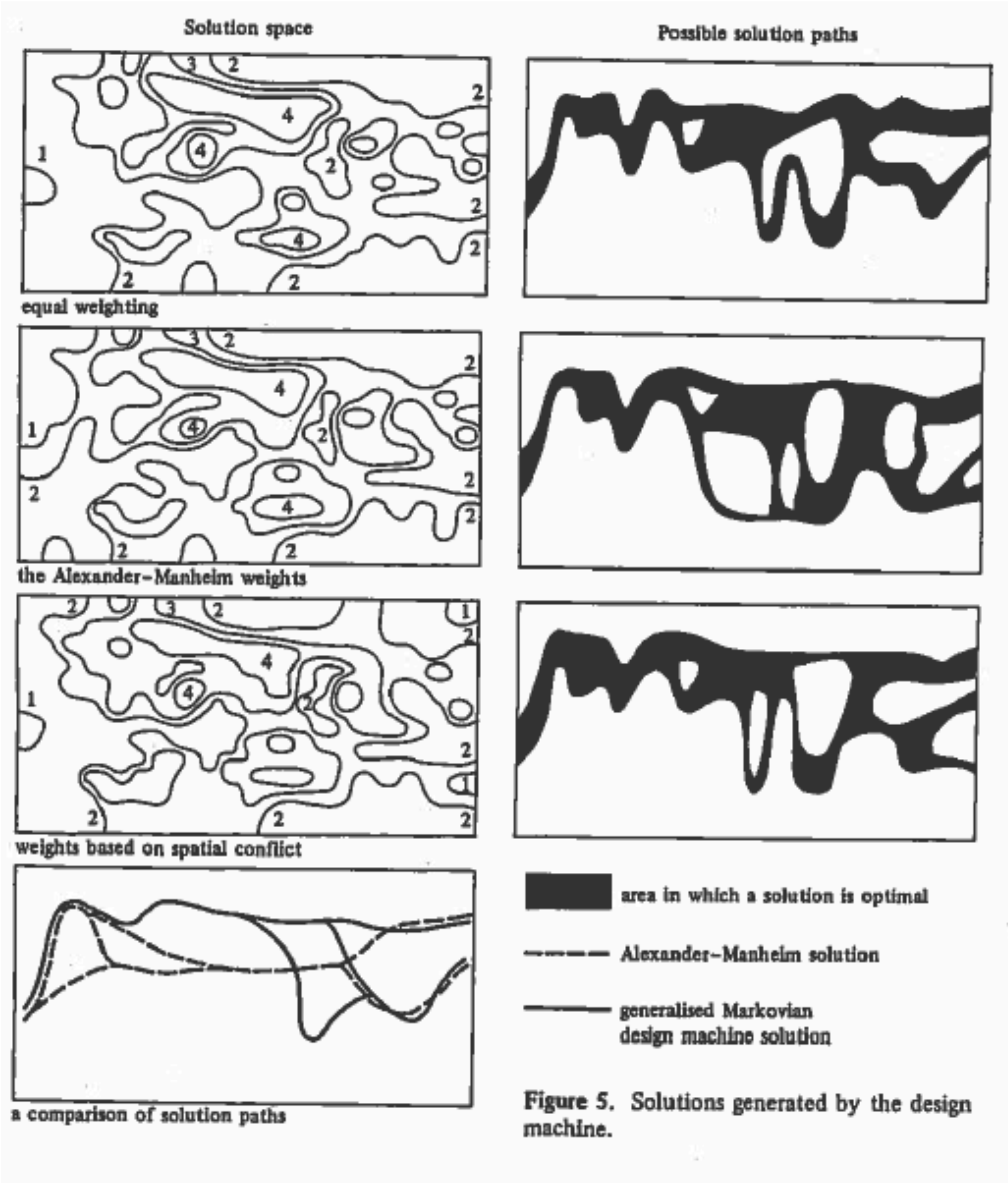


Fig 2.3 *– Batty's Markovian Design machine solutions*

What might a Digital Cadastral Ledger and Block Chain Land Registry Look like?

LAND REGISTRATION[16]

Deeds and titling not only provide critical protection for homebuyers in developed nations—they serve as a basis for investment and economic growth across many developing nations. By securing a unique and non-corruptible record on a blockchain and validating changes to the status of that record across owners, a reliable property record can be created, whether for a piece of land that heretofore had no owner or as a link between stovepiped systems.

Existing pain points:

- License and registry processes are paper-based and fragmented, making transactions costly, inefficient, and vulnerable to tampering
- In the United States, landowners spent $800 million in 2014 and '15 on title insurance to cover risks associated with real estate titles[17]

Blockchain value proposition:

- A decentralized, standardized system for land registration records could reduce the number of intermediaries required, increase trust in identity of transacting parties, increase process efficiencies, and decrease time and cost to process
- Recording property rights via blockchain would enable $2–4 billion in annual cost savings in the United States alone for title insurers through a tamper-proof ledger[18]

Fig 2.4 *– Blockchain Property Ledger – Deloittes and qz.com articles 2017*

Lantmäteriet

Sweden's land registry authority is called the Lantmäteriet. Since last June the body has been testing a way to record property transactions on a blockchain. This could save the Swedish taxpayer over €100 million ($106 million) a year by eliminating paperwork, reducing fraud, and speeding up transactions, according to an estimate by the consultancy Kairos Future, which is also involved in the project.

The Swedish system operates on a private blockchain. This has the land authority and others, like the banks, holding copies of the records. When a land title changes hands, each step of the process is verified and recorded on the blockchain .The system acts as a highly secure and transparent verification and storage service for property transactions

https://chromaway.com/papers/Blockchain_Landregistry_Report_2017.pdf

3 Methodology

3.1 Benchmark Case Studies

Kang et al produced a benchmark dataset and propose a general framework for using CNNs to classify façade building structures from street view images on both regional and city scales of several cities in Canada and the US.

Doersch et al use the same Google Street View repository to use a discriminative clustering approach for automatically finding built environment visual elements that are both frequently occurring and geographically discriminative. This process allows them to develop a “stylistic narrative” which allows for an empirical approach to the qualitative assessment of different geographic locales.

Our study is concerned with applying the general workflow adopted by Kang et al to the street view imagery contained on London’s new build development over the last decade with a view to shining narrative insight into the nature of change to the London’s Urban landscape inspired by the application of Machine Learning Techniques to a geo-informational description of urban place (Doesrch et al – 2015).

3.2 CNNs and Transfer Learning

Convolutional Neural Networks(CNNs) are a variation of a deep neural network(DNN), which in turn are a type of artificial neural network(ANN). The Percepton is the simplest ANN architecture. It is based on a Linear Threshold Unit(LTU) which associates a weight with each of its inputs and applies a sum to the weighted sum of inputs and outputs the result. A Multi Layer Percepton is composed of one or more hidden layers of Layers of LTUs and one final layer called the output layer(an ANN with two or more hidden layers is called a DNN). Every layer except the output layer includes a bias neuron and is fully connected to the next layer. The algorithm computes the gradient of the cost function for each parameter in the network and updates the network parameters with Gradient Descent step(Ageron - 2017).

For each training instance the back propagation algorithm first makes a prediction(forward pass), measures the error, then goes through each layer in reverse to measure the error contribution from each connection, this is known as a reverse pass. Finally it slightly configures the connection weights to reduce the error(Gradient Descent Step). This algorithm, referred to as back propagation, can be used with this logistic function or others (for example, a rectified linear unit or RELU). Back propagation is Gradient descent using reverse mode auto diff (Ageron - 2017).

A CNN mimics the way neurons in the human visual cortex have a small local receptive field, which are tiled together to represent the whole visual field. This biologically inspired and powerful architecture is able to detect all sorts of complex patterns in the any area of the visual field(Ageron - 2017).

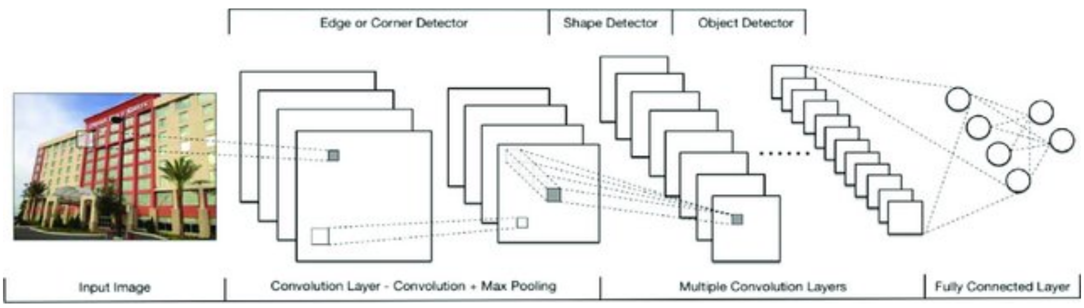


Fig 3.1–*Convolutional Neural Netowrk – Layer Design*

CNNS further use partially connected layers and pooling layers. Unlike a regular DNN the convolutional layers neuron are only connected to pixels in their receptive field. In turn each neuron in the second convolutional layer is only connected to a small rectangle in the in the first layer. This allows the network to concentrate on low level features in the first hidden layer then assemble them into higher level feature in the next hidden layer. Pooling layers, limit the risk of model over fit, by subsampling the input image to reduce the computational load, memory usage and number of parameters Typical CNN architectures stack a convolutional layers then follow with a pooling layer followed by another few convolutional layer stack and pooling layer, and so this process repeats depending on the number of layers

For our research we are using the Inception V3 architecture and it's a lighter version MobileNet, which is easier to run on desktop machinery though at a cost of less accurate overall.

Inception-v3 is trained for the ImageNet Large Visual Recognition Challenge using the data from 2012. This is a standard task in computer vision, where models try to classify images into clases(Szegedy - 2015).

Places 365 is an Scene recognition database holds 205 scene categories and 2.5 millions of images establish new state-of-the-art performances on scene-centric benchmarks (Zhou -2017)

In practice it rare to train a CNN from scratch. Commonly one will use a pre trained CNN either as an initialization or fixed feature extractor. For the former scenario, the classifier is replaced and retrained on to of the CNN on the new dataset of interest and fine tune the pre-trained networks weights by by continuing the back propagation. It is common to see CNN checkpoints of ths network weights used released for public consumption in a pretrained models. Tuning some or more of the layers.

An alternative option is to remove and use last fully connected layer as a fixed feature extractor but trained on a new dataset. Once the vectors for each image that contain the activations of the layer immediately before the classifier, one can train a classifier (e.g. Linear SVM or Softmax classifer) for the new dataset(Karpathy - 2018)

For this study we used the pre-trained Places 365 dataset(via Pytorch) developed using the Caffe Library for image processing and for Tensorflow Lite and Hub Scripts using the tensoflow library.

The top layer receives as input a 2048-dimensional vector (1001-dimensional for Mobilenet) for each image. We train a softmax layer on top of this representation. Assuming the softmax layer contains N labels, this corresponds to learning N + 2048*N (or 1001*N) model parameters corresponding to the learned biases and weights(tensorflow.com– 2018).

It is worth noting that recent benchmark analysis on 13 Transfer Learning classification models reveals that Resnets are the best fixed feature extractors whilst ImageNet accuracy is a good predictor of fine tuning performance(Kornblith et al – 2018).

We have re-used and modified the code and scripts made available from the Tensor flow Online repository, making alterations to allow for running the code on GPUs, timing the train time and for incorporating confusion matrices into the Tensorhub console utility(for measuring training performance)

- **ConvNet as fixed feature extractor**: Here, we will freeze the weights for all of the network except that of the final fully connected layer. This last fully connected layer is replaced with a new one with random weights and only this layer is trained.

- **Transfer learning scenarios**

  Depending on both the size of the new dataset and the similarity of the new dataset to the original dataset, the approach for using transfer learning will be different. Keeping in mind that ConvNet features are more generic in the early layers and more original-dataset specific in the later layers, here are some common rules of thumb for navigating the four major scenarios:

  1. The *target* dataset is **small** and **similar** to the *base* training dataset. Since the target dataset is small, it is not a good idea to fine-tune the ConvNet due to the risk of overfitting. Since the *target* data is similar to the *base* data, we expect higher-level features in the ConvNet to be relevant to this dataset as well. Hence, we:
     - Remove the fully connected layers near the end of the pretrained *base* ConvNet
     - Add a new fully connected layer that matches the number of classes in the *target* dataset
     - Randomize the weights of the new fully connected layer and freeze all the weights from the pre-trained network
     - Train the network to update the weights of the new fully connected layers
  2. The *target* dataset is **large** and **similar** to the *base* training dataset. Since the *target* dataset is large, we have more confidence that we won't overfit if we try to fine-tune through the full network. Therefore, we:
     - Remove the last fully connected layer and replace with the layer matching the number of classes in the *target* dataset
     - Randomly initialize the weights in the new fully connected layer
     - Initialize the rest of the weights using the pre-trained weights, i.e., unfreeze the layers of the pre-trained network
     - Retrain the entire neural network
  3. The *target* dataset is **small** and **different** from the *base* training dataset.
     Since the data is small, overfitting is a concern. Hence, we train only the linear layers. But as the *target* dataset is very different from the *base*dataset, the higher level features in the ConvNet would not be of any relevance to the *target* dataset. So, the new network will only

use the lower level features of the *base* ConvNet. To implement this scheme, we:

   - Remove most of the pre-trained layers near the beginning of the ConvNet
   - Add to the remaining pre-trained layers new fully connected layers that match the number of classes in the new dataset
   - Randomize the weights of the new fully connected layers and freeze all the weights from the pre-trained network
   - Train the network to update the weights of the new fully connected layers

4. The *target* dataset is large and different from the *base* training dataset. As the *target* dataset is large and different from the *base* dataset, we can train the ConvNet from scratch. However, in practice, it is beneficial to initialize the weights from the pre-trained network and fine-tune them as it might make the training faster. In this condition, the implementation is the same as in case 3.

(Karpathy -2018)

New dataset is small but very different from the original dataset. Since the data is small, it is likely best to only train a linear classifier. Since the dataset is very different, it might not be best to train the classifier form the top of the network, which contains more dataset-specific features. Instead, it might work better to train the SVM classifier from activations somewhere earlier in the network.

New dataset is small and similar to original dataset. Since the data is small, it is not a good idea to fine-tune the ConvNet due to overfitting concerns. Since the data is similar to the original data, we expect higher-level features in the ConvNet to be relevant to this dataset as well. Hence, the best idea might be to train a linear classifier on the CNN codes. (Tensorflow Org, Online Documentation -2018)

Softmax Classifiers Explained

The Softmax classifier is a generalization of the binary form of Logistic Regression. Just like in hinge loss or squared hinge loss, our mapping function f is defined such that it takes an input set of data x and maps them to the output class labels via a simple (linear) dot product of the data x and weight matrix W:

$$f(x_i, W) = Wx_i$$

However, unlike hinge loss, we interpret these scores as unnormalized log probabilities for each class label — this amounts to swapping out our hinge loss function with cross-entropy loss:

$$L_i = -log(e^{s_{y_i}} / \sum_j e^{s_j})$$

(Rosebrock - 2019)

3.3 Context: Picturing the scene

Ground Truth Verification plays an important role in the Image Classification process and in work involving Buildings and Addresses. We begin by establishing a strong Visual Understanding of the data and subject in question and build on this approach throughout the study.

Quick Look at Quantity and Types of Building Data

There are upwards of 65000 developments recorded in the LDD with Croydon and Westminster delivering the greatest number. As the LDD also records, for example, Permissions and Demolitions this figure does not necessarily reflect the boroughs that have delivered the largest number of new units or floor space built.

London New Build: What? Where? When?

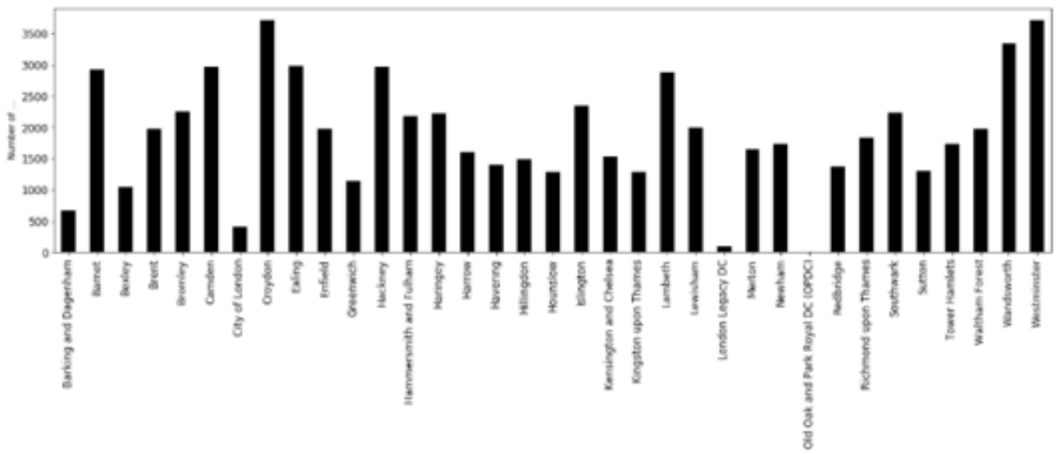


Fig 3.2 *–London New Developmnet Since 2004 by Borough*

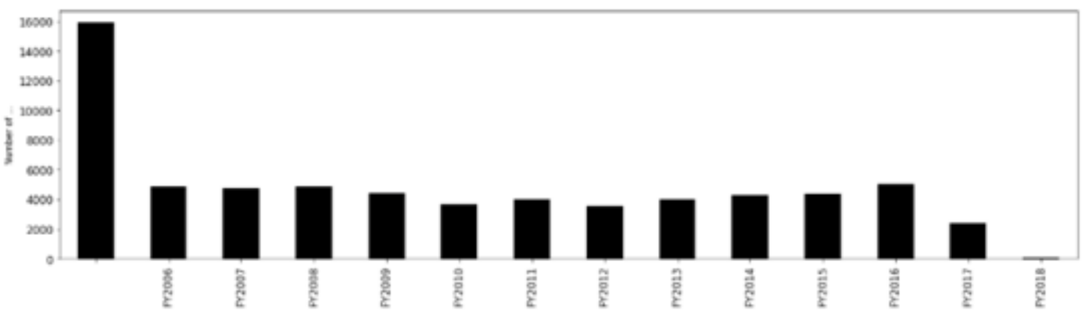
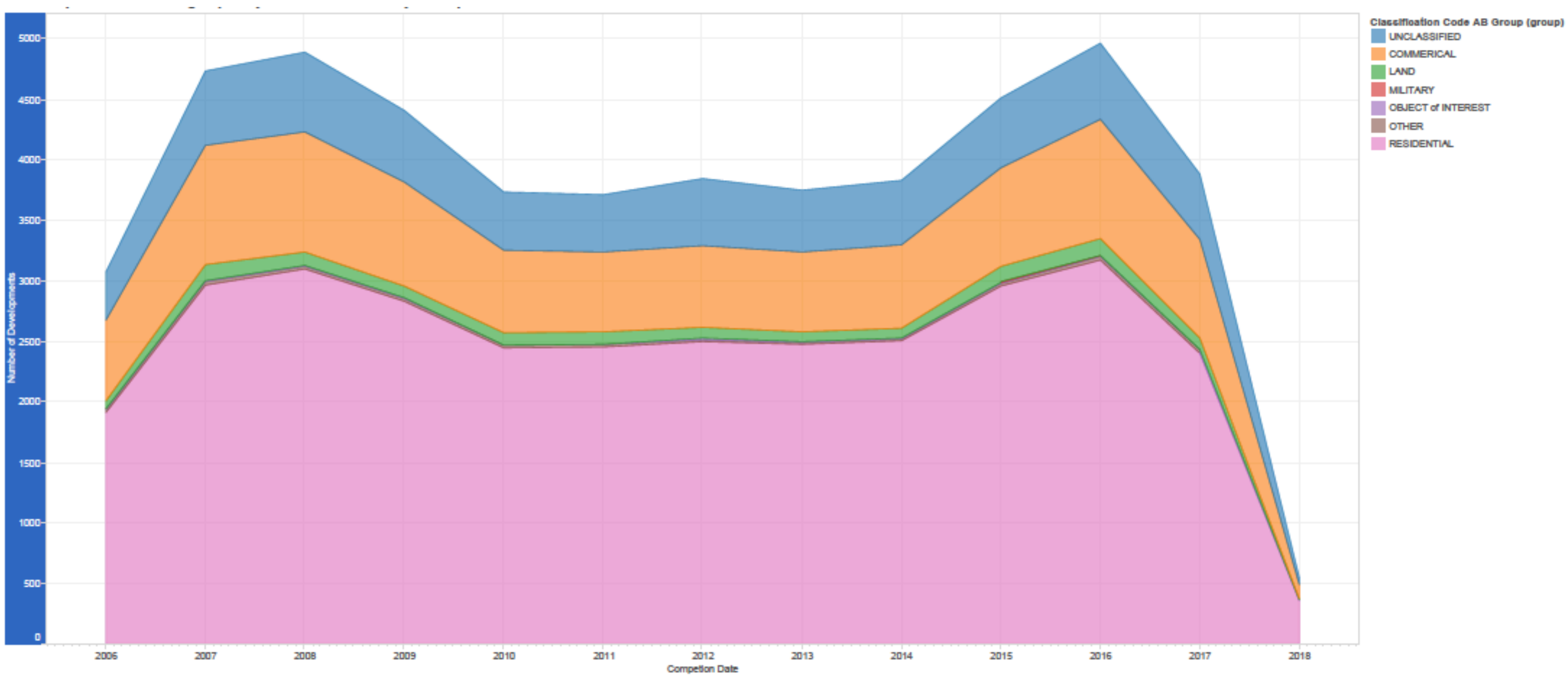


Fig 3.3 *– Completions vs Permissions (The first column represents Building Non Completions and demolitions)*



LDD Completions overtime grouped by Address Base Primary Description

Fig 3.4 *–London New Development By Use Class.*

The 2 spikes reflect the 2008 recession and post recession recovery periods as well as the time lag frequently found in recording building completions.

3.4 Data Model

The datasets used for this study were the London Development Database (LDD) maintained by the Greater London Assembly(GLA), the Ordnance Survey AddressBase Premium dataset and the Google StreetView Image repository:

LDD	**Address Base**	**Google StreetView**
Location	Address	Building Façade Image
Development Description	Property Classification	Image Meta Data
Density(Number of Units, Area Size)		
Geo Reference	Geo Reference	Geo Reference

The LDD monitors new building completions across London since approx. 2004 and includes planning consents that include any new loss or gain of residential units, major changes of use (e.g. 1000m2 or creation of > 7 bedrooms)), and the loss or gain or change of use of open .

The address base Classifications are sourced from the Local Land and Property Gazetteers and from Ordnance Survey large-scale data and include Primary (e.g. Commercial), Secondary(Hotel) and Tertiary(Youth Hostel) descriptions that increase in finer grained detail:

C	Commercial	**H**	Hotel/Motel/ Boarding/ Guest House	**01**	Boarding/Guest House/ Bed And Breakfast/Youth Hostel	Object defined by local government contributing authority, includes: Commercial Lodging.	**YH**	Youth Hostel

Typically a Completions entry on the LDD register will record the X and Y Coordinates(in a British National Grid projection) as generated as part of the validation stage of the UK Planning application life cycle. For Changes of Use or Minor Developments this will generally be adequate for a geo referenced match, however the Land Parcel will frequently not have gone through the standard Street Name and Numbering process yet, and been allocated an official, validated address and postcode. On Large developments the matching problem is exacerbated and over reliance on the geo coordinates as presented in the LDD will reap many mismatches.

To join the validated AddressBase data to the LDD, a buffer spatial join was carried out on the address Base premium data. Only the top matching address was returned.

We sought the assistance of by the Tower Hamlets Council GIS and SNN Team in procuring the licensed AddressBase dataset and in carrying out the spatial join of the selected London Non New Developmnet and New Development Addresses (n = 500k) using ESRI SDE. The buffer join task using ArcEngine v10.7 took over 5 hours to complete.

3.5 Auxiliary Visual Descriptor Data

Transport for London's (TFL) Public Transport Accessibility Levels (PTALs) London dataset and CACI Ltd’s Acorn geo-demographic classification dataset were also referred to as auxiliary datasets during the model build stage providing socio-economic, geographic and visual context as well as assistance in ground truth verification in the image classification stages.

Place Pulse is a crowdsourcing effort to map urban perception. By asking users to select images from a pair, Place Pulse collects the data needed to evaluate people's perceptions of urban environments.

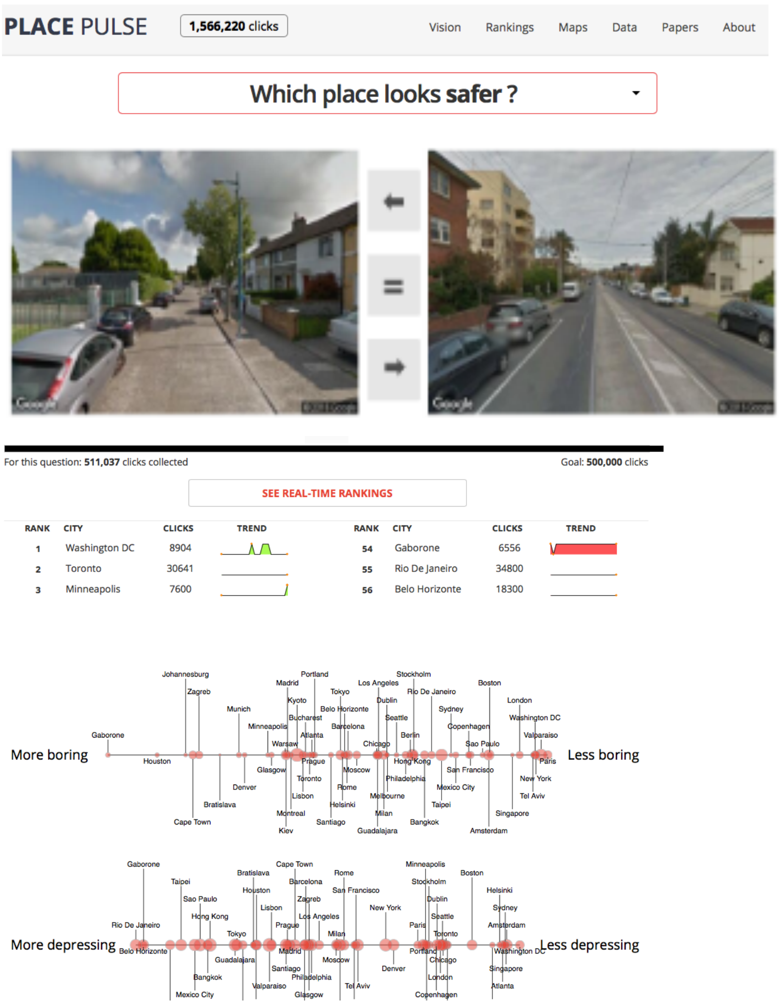


Fig 3.5 –MIT Place Pulse Urban Perception Data Project. We applied their Training Dataset on London New Build during the Data Model Exploration Stage

Transport for London's (TFL) Public Transport Accessibility Levels (PTALs)
Public Transport Accessibility Levels(PTAL) are a detailed and accurate measure of the accessibility of a point to the public transport network, taking into account walk access time and service availability. The method is essentially a way of measuring the density of the public transport network at any location within Greater London. Each area is graded between 0 and 6b, where a score of 0 is very poor access to public transport, and 6b is excellent access to public transport.

Legend:

C = Commercial
R = Residential

L, M, O, P, X, Z = Other Address Primary Address Types

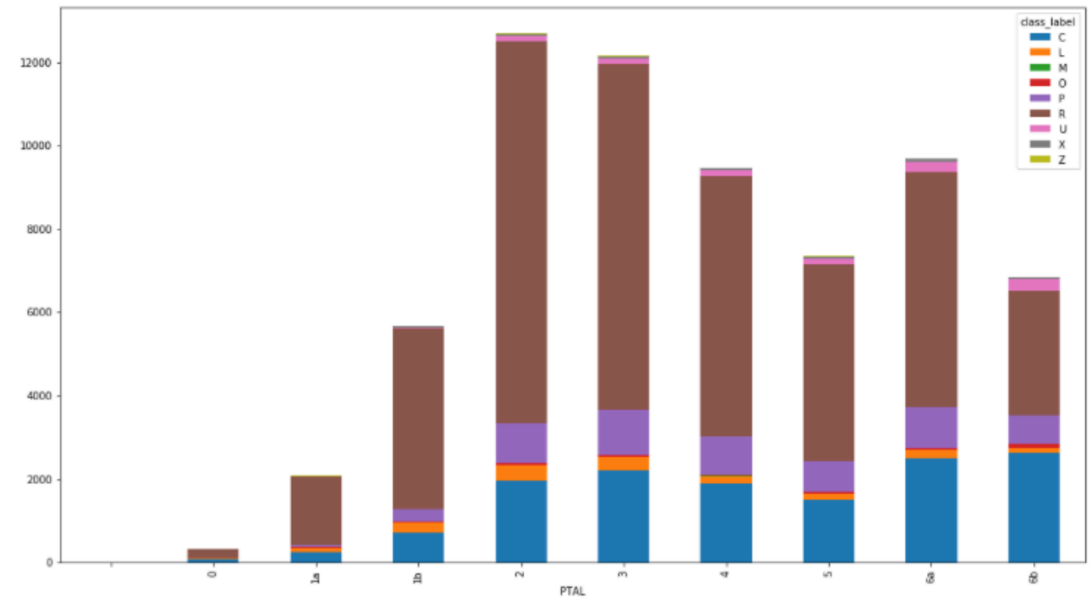


***Fig 3.6** –London Building Development since 2004 by PTAL. Greater access to Public Transport is accompanied by a more equal ratio of Commerical and Residential Address Types reflecting inner and suburban urban areas.*

Acorn is a consumer classification that segments the UK population. By analyzing demographic data, social factors, population and consumer behavior, it provides precise information and an understanding of different socio economic categories of people.

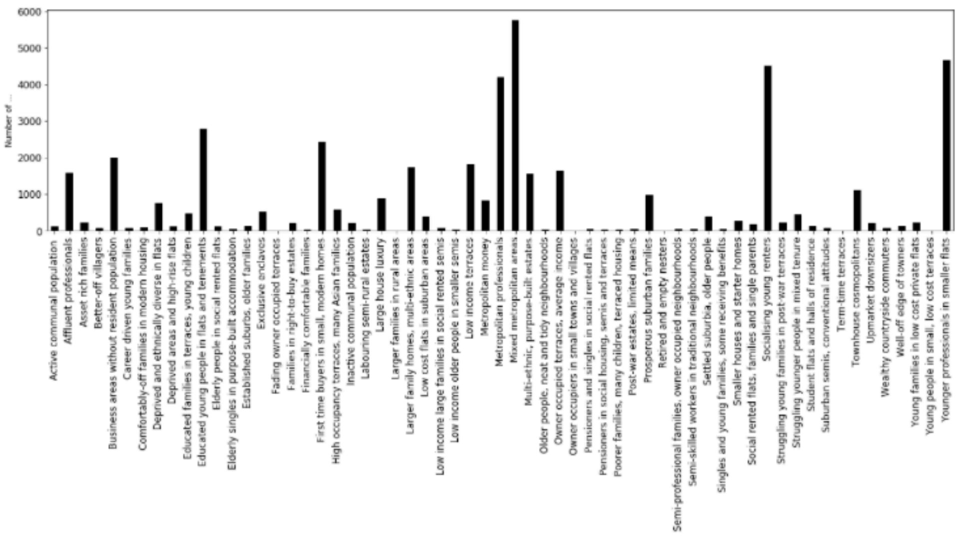
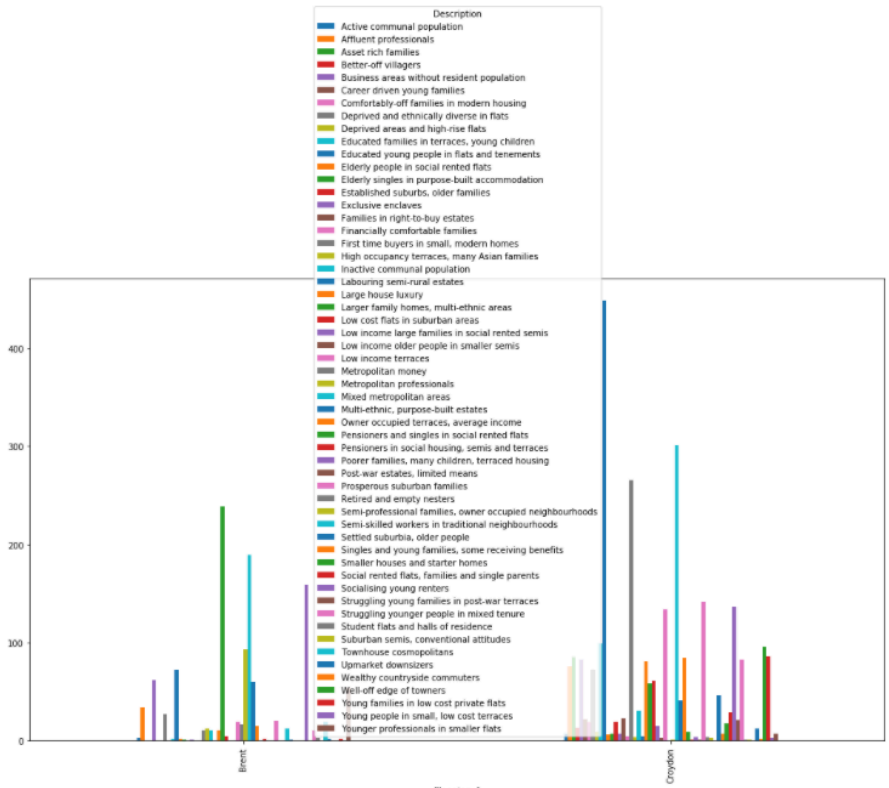


***Fig 3.7a amd 3.7 b** –London New Development by CACI Geodemographic Class. The largest number of development come under the Metropolitan and mixed Metropolitan demographic classes. Croydon leads on "Educated Young People in Flats and Tenaments". Brent leads on "Larger Family Homes, Multi Ethnic"*



3.6 Research Limitations/Challenges

The nature of the Building Completion lifecycle(e.g. phased large scale developments which can take a long time to be recorded) and the collection of its data(collected and submitted by Local Authorities) can be a slow process.

For our requirement of needing to obtain enough data(valid building completions and streetview images) to train an ML Classification model(>100 Images per Category but ideally 1000+ per Label), we zoned in on the 10 year period in and around 2006 and 2016.

We found that Google Street View images were in the greatest and most reliable supply(in terms of having actual images of the property in question) dating back several years. We therefore filtered the data model accordingly so that we looked at permissions that had already completed and that had Street View Images readily available.

The recurrent issue in face of being able to run a train a model on data labels that were differentiated enough for the CNN to be effective was the issue of how to deal with anomalies which add noise to the feature space being examined.

These scenarios underline the importance of the processing stage of the workflow, but also in the importance of being able to draw clear enough boundaries between the visual categories we are trying to differentiate and identify. This is a process of obtaining a clear defined pattern or signal that can be separated from randomness or noise.

The significance of the decision boundary process that is ultimately selected to draw out the boundaries of interest, brings to mind the perenial Modifiable area unit problem(MAUP) but also alludes to the general Qualitative vs. Quantitative division which highlights the wide diversity and variety in effect and outcome that is dependent on the method selected for any given case study.

Address Matching is commonly faced challenge and in appraoching our research we needed to deploy a means of verfying or double checking the georefencing component to be confident that the ML Classification stages were or were not functioning correctly.

The already mentioned problem of labeling anomalies provided a further considerable challenge in the form of building object occlusions and inaccessible classification targets.

Maintaining immutable geo refrencing for these objects across data sets spanning over time in a rapidly changing built envinroment is a further challenge that further emphasies the need for a prominent validation component to the research.

3.7 Data Filtering

The raw data model(new development plots coupled with a street level image record and additional descriptor data joins) yields approximately 70000 records and at first sight appears to offer us a rich seam of London new build examples on a scale not far off the Kang et Al study. However upon further consideration it is apparent that several factors will necessitate the removal of a large number of these records.

Figure 3.10 reveals familiar and, for the building type classification task at hand, pertinent primary categories of building such as the Commercial and Residential Primary AddressBase Descriptor classes but the chart also highlights numerous building type labels that we will need to discard(unclassified, parent shell, mixed use) as they do not provide us with a sufficiently distinct visual category of building form.

Orphaned Use Class categories which do not belong to a discernable class of building(e.g. a point of interest could be a member of several address categories) will also be discarded. In addition, certain LDD Record types provide us with visually confusing and ambiguous street level image records: incomplete building permissions and demolitions.

```
- 8090 House Images
- 16814 Flat Images
- 1412 Office IMages
- 2789 Retails Images
- 549 Industrial Images
```

Fig 3.8 *– Population Data Size after Data Filtering*

After Data Filtering we have a significantly reduced dataset of London new build delivered since 2004 however we still appear to have an adequate amount of data for a machine learning exercise.

We now need to proceed to looking at whether this translates into an even spread of property types and what the quality of this address and image data is like.

3.8 Creating Machine Learner Friendly CNN Classification Label Sets

Our Image classification process is concerned with Building Appearance and form. Quantitative Building and Address Data Typologies tend to be based on Building Usage. For our Machine Learning exercise we will need to employ a process that projects the data descriptive categories into a discernable and distinct visual physical type.

Figure 3.9 describes the primary grouping of the OS AddressBase types(generally represented by the first letter). Fig 3.10 illustrates is how the distribution of London New Build looks when it is divided up into its address type sub hierarchies.

More detailed explanation of the AddressBase Schema is provided later on and in the Label Set Creation Workflow Stage.

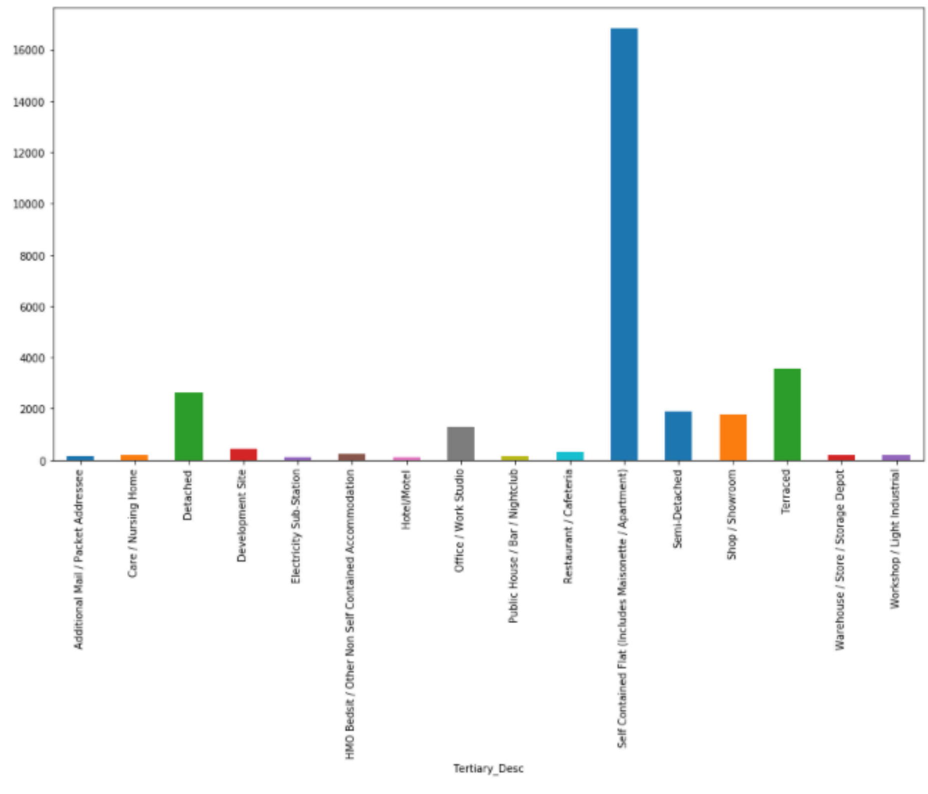


FIg 3.9 *– London Development by Teritary Addressbase Class*

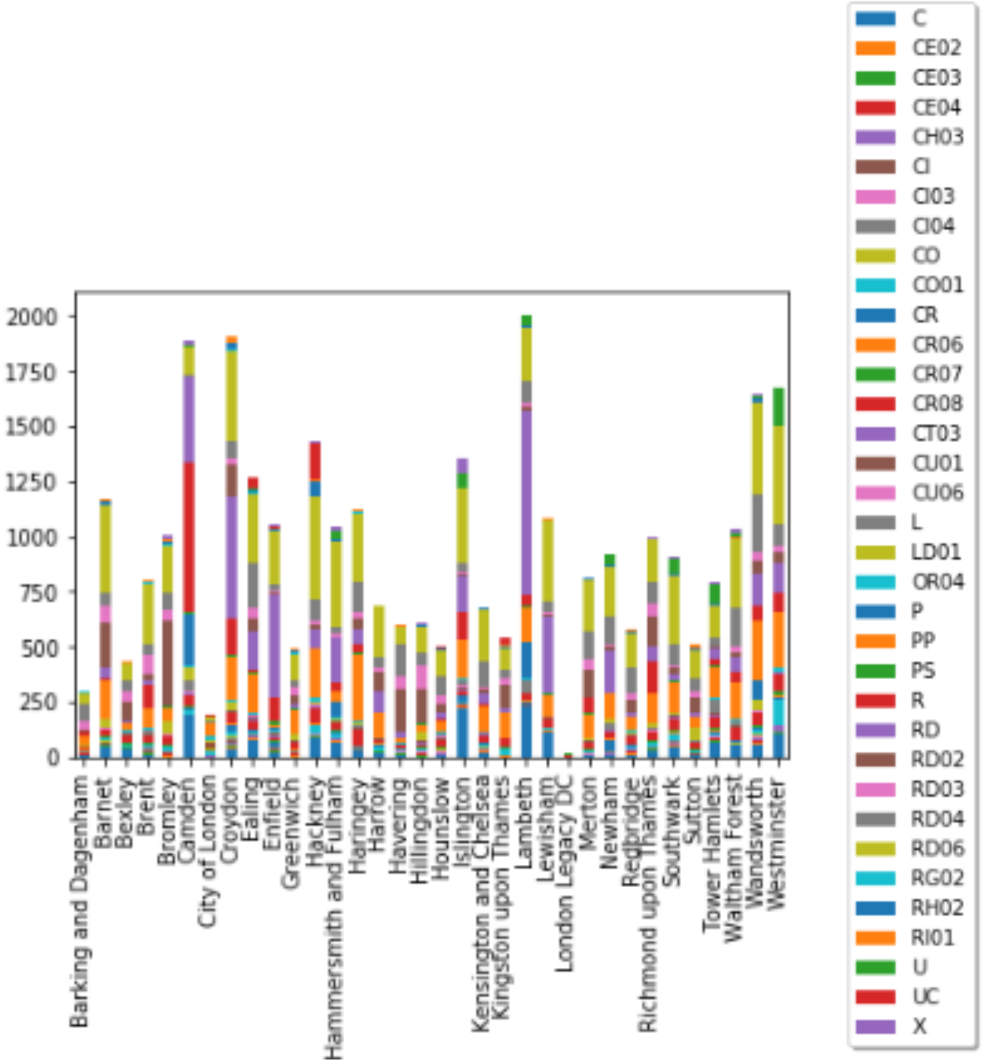


***Fig 3.10** – London Development by Primary Addressbase Class*

Primary code	Primary description
C	Commercial Attracts Non Domestic Rates and/or use is of a business nature
L	Land
M	Military Military Defence Site
O	Other (Ordnance Survey only)
P	Parent Shell
R	Residential
U	Unclassified
X	Dual Use
Z	Object of Interest

***Fig 3.11** –AddressBase Premium Descriptor Class*

The above table is noisy and hard for humans to visually process. The self same issue holds for a machine learner classifier. The challenge is how to focus on the properties of interest to clarify our motivation and tailor it to the method based outcome. In our case this is Image Classification and Prediction.

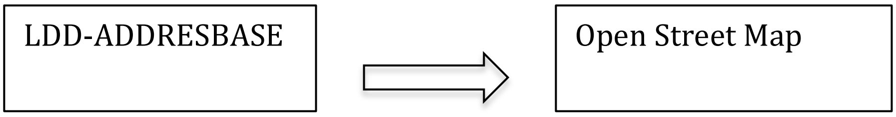
We can see that using the AddressBase categories generates a large number of categories. Whilst pre trained CNNs typical perform better on labels that are sub divided(see Tensorflow technical notes in the CNN training workflow Stage section) our initial label runs uncovered that the wide variety and mix of uses did not translate into separate physical or visual objects.

Initial trial Image Classification runs were attempted with the LDD and AddressBase Premium Property Types. The results scored low in accuracy. Similar Trial Runs were almost made with the appended Acorn, PTAL , Place Pulse Auxiliary Property Classifiers. These were also unsatisfactory.

CNN training tests were also carried out on Use Class categories for properties in Tower Hamlets. Use Class data is typically maintained by Local Authority Planning Departments for carrying out analysis and inventory in the Planning process. Once again the data was able to provide a sufficiently accurate Image Classification result.

From these trial runs and from our analysis of the spread of property type classes it was apparent our Neural Network would need to be sub divided into more primitive types.

3.9 Open Street Map to the Rescue



The Open Street Map Building Types Classification methodology features a map based schema element that focuses on objects that relate to building and architectural form. Kang et al successfully applied this translation technique in their ML work on US Cities. Consequently we sought to find a way to map and translate the Address Base to OSM Types for our London based dataset.

An initial export of both datasets into SQL Server 2014 and with a heuristic approximate mapping translation of AB Tertiary Types provides us with 12000 houses and 21000 apartments.

Fitting the LDD data into these categories is not a smooth fit. In the visualization below we highlight the number of records that fit into these categories but also highlight the number of records that re left orphaned:

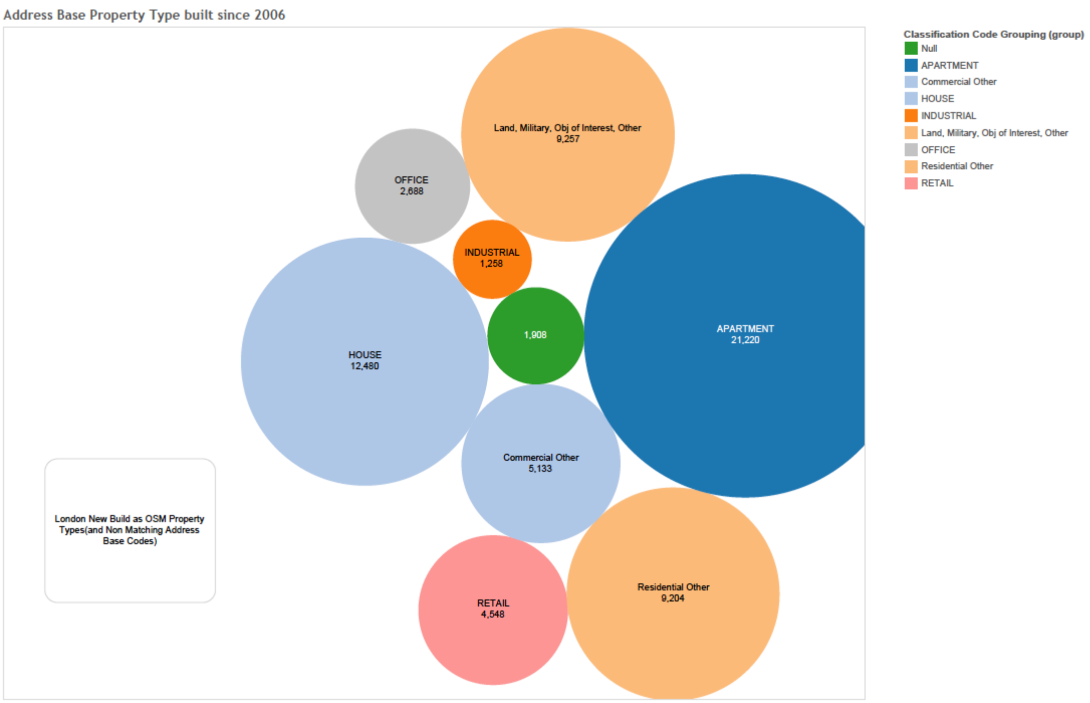


***Fig 3.12** –London Development by Open Street Map Use Class*

For our initial exploration of the data and to find categories that are able to discern different building objects and ignore mutual use types membership, we matched relevant AddressBase Types with the Open Street Map Categories of Interest(table 3.13).

The next step would be in finding a more satisfactory method of translating Address Base codes into OSM Building types.

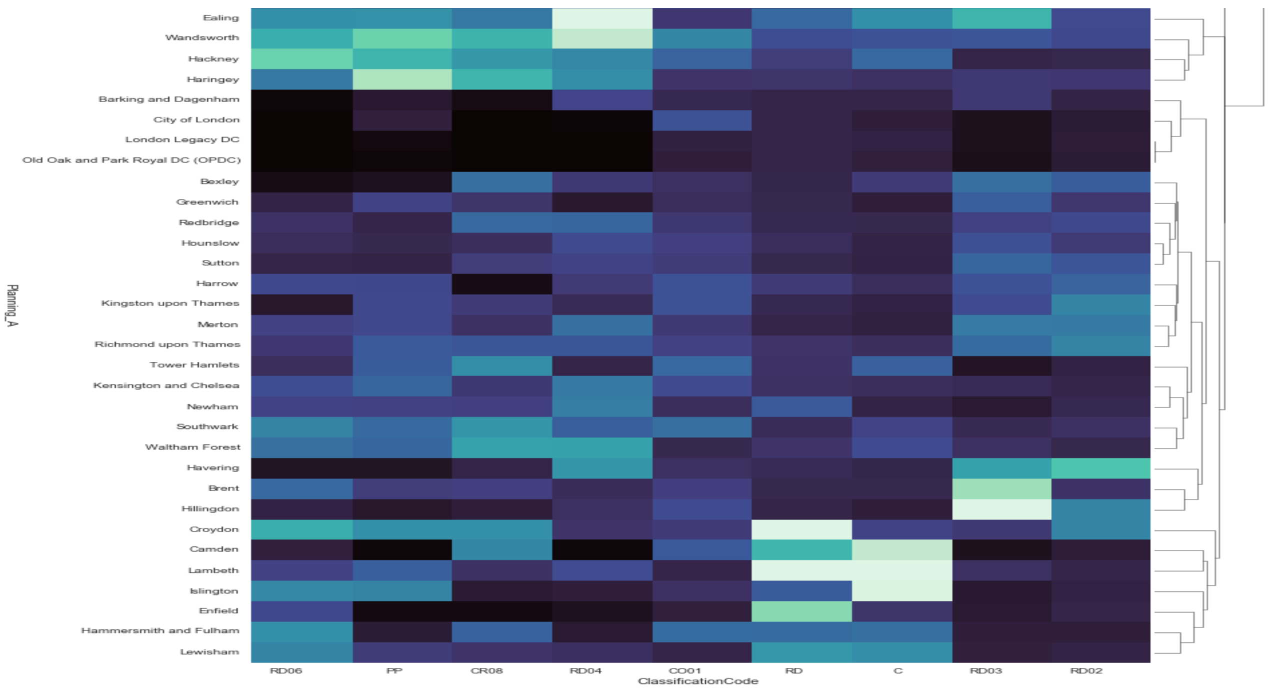
OSM Building Types	
Church	*A building that was built as a church*
Garage	*A building suitable for the storage of one or possibly more motor vehicle or similar*
House	*A dwelling unit inhabited by a single household (a family or small group sharing facilities such as a kitchen)*
Industrial	*A building where some industrial process takes place*
Office building	*A building where non-specific commercial activities take place*
Retail	*A building primarily used for selling goods that are sold to the public*
Roof	*A structure that consists of a roof with open sides, such as a rain shelter, and also gas stations*
Apartment	*A building arranged into individual dwellings, often on separate floors. May also have retail outlets on the ground floor.*

Fig 3.13 *–OSM Categories of interest*

Address Base Types
Detached
SemiDetached
Terraced
Self Contained Flat
HMO
Communal Residence
Commercial
Industrial

Fig 3.14 *–Equivalent AddressBase Types*

Fig 3.15 *–Heatmap of London Development by AddressBase Types. Darker patches represent large numbers, Light patches smaller numbers. Larger amounts of RD06, RD04 and CR08 can be seen in Westminster, Barnet, Ealing, Wandsworth*



Code Descriptions:

RD06 = Flat
PP = N/A
CR08 = Retail
RD04 = Terraces
C001 = Office
RD = N/A
C = Commercial
RD03 = Semi- Detached
RD02 = Detached

3.9 Crisp Sets and Fuzzy Sets

A traditional set is a dichotomous object with set membership being a binary condition, a member being either in or out. Fuzzy sets retain the states of full and non membership but also allow for an interval fuzzy membership score depicting degrees of membership(Ragin –2000).

When faced with the dilemma of defining categories or typologies of interest, 2 considerations relating to generalizability typically are cited.

Necessity of cause determines whether a factor is always required to effect an outcome. *Sufficiency* determines whether the factor on its own is capable of producing the outcome in question. Fuzzy sets and their use of configuration and controls to grade a degree of membership(for instance providing a cross over point of > .5) allow for a systematic but pragmatic methodology to keep an empirical enquiry moving in the face of a lack of undisputed and dichotomous data.

(Hullermeier -2011) details how Fuzzy Set Theory approaches can be used in Machine Learning and data mining with fuzzy information granulation being an ideal tool for trading off accuracy against complexity and understandability.

As we shall see the problem of defining boundaries between property types and uses is a problem for a Machine Learning Classification task and for approaching urban typologies in general. Varying definitions of a building exist and are used according to the purpose of the descriptive task.

Use Classes are used in planning. One building can several use classes. Especially in the current climate where density in terms of housing need and availability is a major concern in the London Planning process.

Address base use classification code which deals with Local Authority submitted address data to the National Gazetteer hub.

Open Street Map in addition has layer data on building types although this data is limited to a uninformative yes or no category in many instances in the UK data.

A challenge for our study is in selecting a typology or family of labels robust enough for carrying out a Classification exercise. This process resonates with on the motivation of uncovering of Geo Informative narrative through computational urban geography and will ultimately go onto frame our final analysis and conclusion.

3.10 Fuzzy Set Approach for deriving a CNN Friendly Building Type Nomenclature

Using simple logical statements to encapsulate the predominant groupings, we arrive at the following:

Terraced Building = Terraced Building + Shop Use • Terraced Building + Minor Flat Use • Terraced Building + House Use

Apartment Building = Multi Storied Building + Shop Use • Multi Storied Building + Minor Flat Use • Multi Storied Building + Office Use

House Building = House Building + Shop Use • House Building + Minor Flat Use • House Building + House Use • House Building + Office Use • House Building + (Minor) Use

Industrial Building = Site/Park Object + Office Use • Site/Park Object + Retail Use • Site/Park Object + Mixed Use(e.g. Residential)

Retail Building = Retail Building + Mixed Use • Retail Building + Residential Use • Retail Building + Office Use

Office Building = Retail Building + Mixed Use • Retail Building + Residential Use • Retail Building + Office Use

In this statement multiplication(denoted with midlevel dots) indicates combinations of characteristics(logical *and*); addition indicates alternate combinations(logical *or*). Translated, these equation state the multiplicity of mixed use that can be found in a single building artifice type, and allow us the possibility of methodically approaching the differences and nuance.

Clearly Apartment and Office blocks share a great deal of attributes however it is equally fair to say we are dealing with clearly delineated distinct objects from the design/ aesthetic and also ML Classification task point of view.

If the variety in Use and Purpose of a building is not ostensibly apparent in physical form, it is possible to see where a machine learning algorithm(dependent on extracting features such as edge, lines and texture) will have difficulty in an Image Classification Task.

A key issue in obtaining precise and meaningful results will depend on how we configure the boundaries between our image categories.

__Fig 3.16__ –Fuzzy Controls for AB to OSN Address Type Mapping

OSM Property Types (CNN Friendly): APARTMENT HOUSE INDUSTRIAL RETAIL NO OFFICE REASON: OFFICE look like APARTMENT or RETAIL	OSM Property Types (CNN Friendly): APARTMENT HOUSE INDUSTRIAL OFFICE NO RETAIL REASON: RETAIL look like APARTMENT or OFFICE	OSM Property Types (CNN Friendly): APARTMENT HOUSE INDUSTRIAL MAJOR OFFICE MAJOR RETAIL	OSM Property Types (CNN Friendly): APARTMENT HOUSE INDUSTRIAL MAJOR OFFICE MAJOR RETAIL SMALL OFFICE SMALL RETAIL
FUZZY SOLUTION: Combine RD2+3+4 Filter Out RD06 < 10 Exclude OFFICE	FUZZY SOLUTION: Combine RD2+3+4 Filter Out RD06 < 10 Exclude RETAIL	FUZZY SOLUTION: Combine RD2+3+4 Filter Out RD06 > 10 Filter Out OFFICE < .08 Filter Out RETAIL < .08	FUZZY SOLUTION: Combine RD2+3+4 Filter Out RD06 > 10 Filter Out OFFICE < .08 Filter Out RETAIL < .08 New Class for

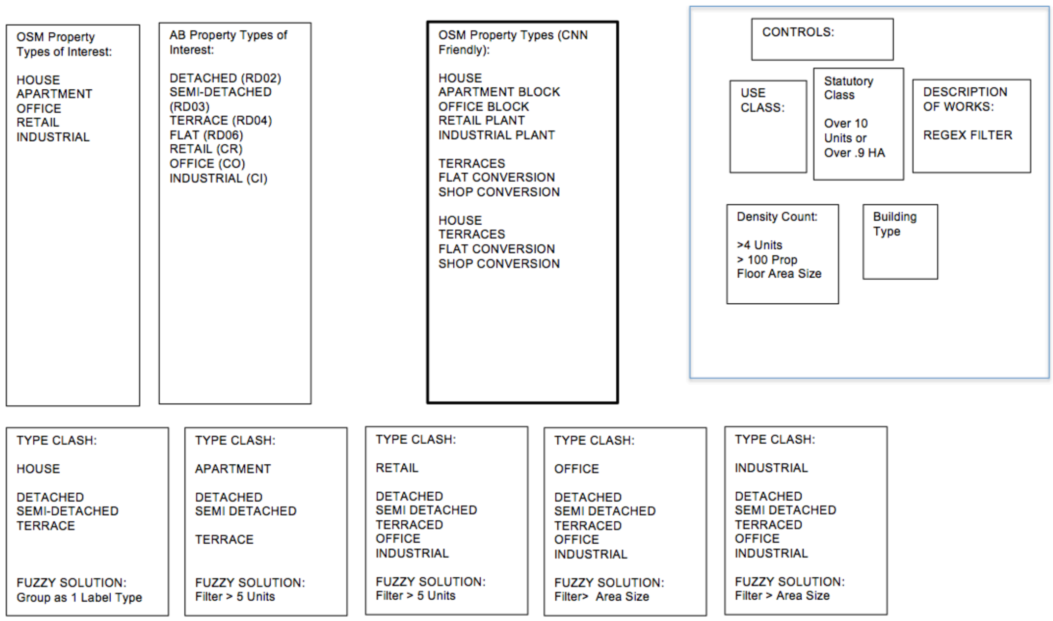


***Fig 3.16** – Fuzzy Controls for AB to OSN Address Type Mapping*

Unit numbers and Floor Area were the chosen filters or control used to delineate meaningful labels for the CNN Trainer.

Offices and Retail had distinctive design qualities(business like building facades and loud signage, respectively). Flat conversion which would appear in houses, high street terraces alike would prove to be the most problematic.

Detached, Semi Detached and Terraced houses had to be grouped together as the CNN found it hard to differentiate between these 3 Addbase categories.
For flats we chose a cut off point of 4 units. This didn't match the 10 unit limit that defines the UK Statutory Classes for Major and minor development, However in our early test we found this had the most satisfactory results for the CNN to differentiate between the classes. We also focused on large floor space to for industry and offices.

In essence we wanted to label all London Buildings in terms APARTMENT BLOCKS, OFFICE BLOCKS, INDUSTRY PARK, RETAIL OUTLET and HOUSES.

3.10 Technical Platform, Tools and APIs Used

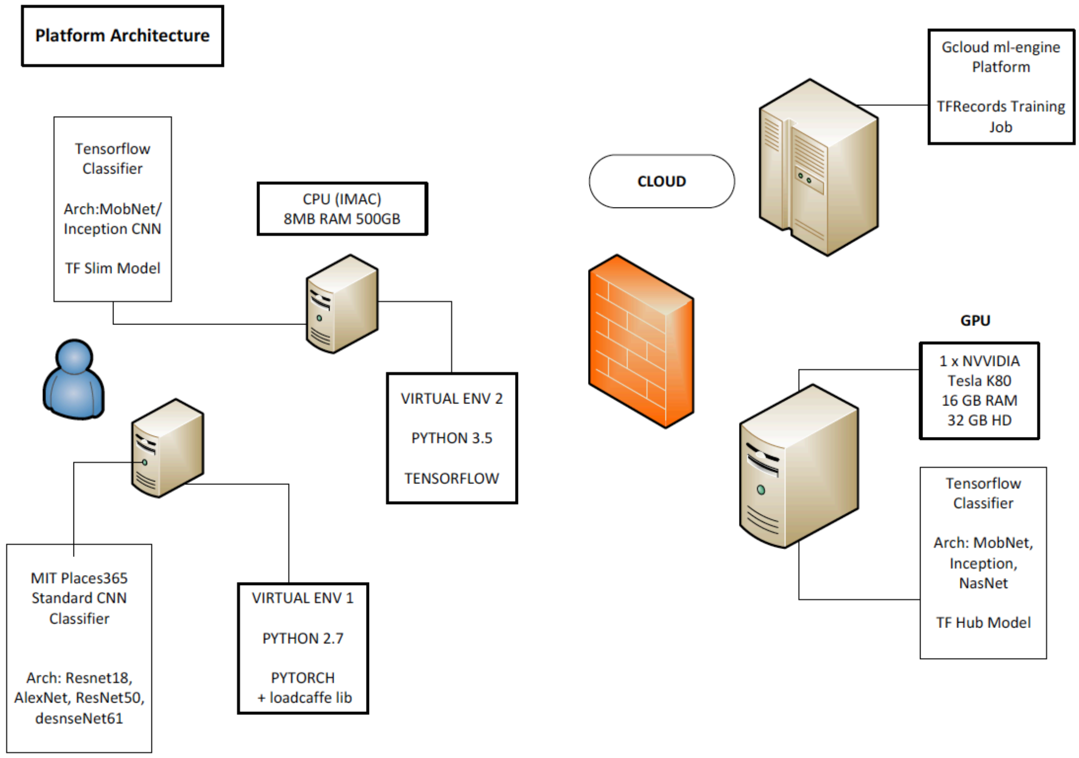


***Fig 3.17** –Workflow Technical Architecture*

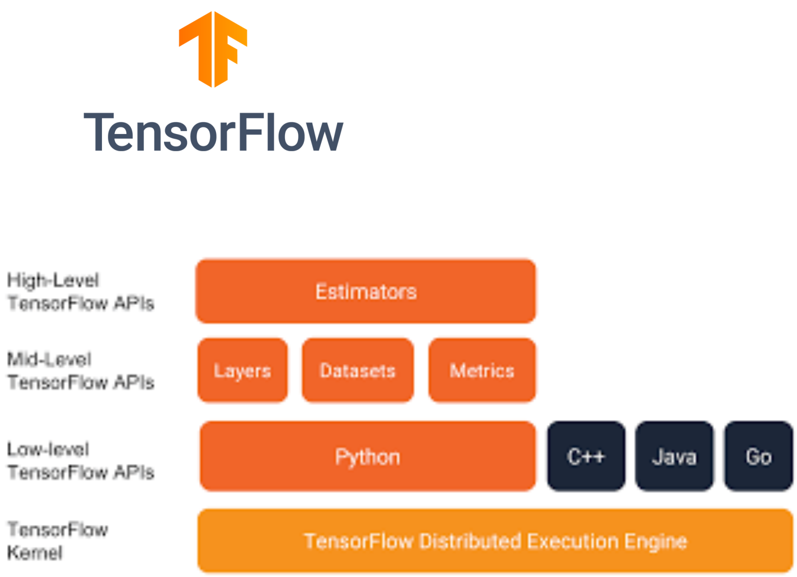


***Fig 3.6** –Tensorflow Platform*

3.11 Tensorflow Platform

2 versions of the Tensorflow scripts were used through out the project. One implentation uses the Tensorflow Slim version of the tensorflow library resource, the other harnesses the Tensorhub repository.

Both scripts take a Inception v3 or Mobilenet model trained on ImageNet images, and train a new top layer that can recognize other image classes. Other model architectures were also used in pursuit of better accuracy and general comparison.

The Inception v3 top layer receives as input a 2048-dimensional vector (1001-dimensional for Mobilenet) for each image. It trains a softmax layer on top of this representation. Assuming the softmax layer contains N labels, this corresponds to learning N + 2048*N (or 1001*N) model parameters corresponding to the learned biases and weights()ensoflow.com- 2015).

This release contains the model definition for MobileNets in TensorFlow using TF-Slim, as well as 16 pre-trained ImageNet classification checkpoints for use in mobile projects of all sizes. The models can be run efficiently on mobile devices with TensorFlow Mobile.

Model Checkpoint	Million MACs	Million Parameters	Top-1 Accuracy	Top-5 Accuracy
MobileNet_v1_1.0_224	569	4.24	70.7	89.5
MobileNet_v1_1.0_192	418	4.24	69.3	88.9
MobileNet_v1_1.0_160	291	4.24	67.2	87.5
MobileNet_v1_1.0_128	186	4.24	64.1	85.3
MobileNet_v1_0.75_224	317	2.59	68.4	88.2
MobileNet_v1_0.75_192	233	2.59	67.4	87.3
MobileNet_v1_0.75_160	162	2.59	65.2	86.1
MobileNet_v1_0.75_128	104	2.59	61.8	83.6
MobileNet_v1_0.50_224	150	1.34	64.0	85.4
MobileNet_v1_0.50_192	110	1.34	62.1	84.0
MobileNet_v1_0.50_160	77	1.34	59.9	82.5
MobileNet_v1_0.50_128	49	1.34	56.2	79.6
MobileNet_v1_0.25_224	41	0.47	50.6	75.0
MobileNet_v1_0.25_192	34	0.47	49.0	73.6
MobileNet_v1_0.25_160	21	0.47	46.0	70.7
MobileNet_v1_0.25_128	14	0.47	41.3	66.2

Choose the right MobileNet model to fit your latency and size budget. The size of the network in memory and on disk is proportional to the number of parameters. The latency and power usage of the network scales with the number of Multiply-Accumulates (MACs) which measures the number of fused Multiplication and Addition operations. Top-1 and Top-5 accuracies are measured on the ILSVRC dataset.

***Fig 3.19** –Tensoflow Slim Mobile Net Model Variations*

PyTorch

The Pre processing stage used allowed us the opportunity to use an alternative Machine Learning implementation library(PyTorch).

Tensorflow is based on Theano and has been developed by Google, whereas PyTorch is based on Torch and has been developed by Facebook.

One main difference between the two is the way these frameworks define computational graphs. While Tensorflow creates a static graph, PyTorch utilizes a dynamic graph approach. In Tensorflow, you first have to define the entire computation graph of the model and then run your ML model. But in PyTorch, you can define/manipulate your graph on-the-go(note: TF 2.0 and the TF Keras Implementation introduce eager execution, which addresses this difference).

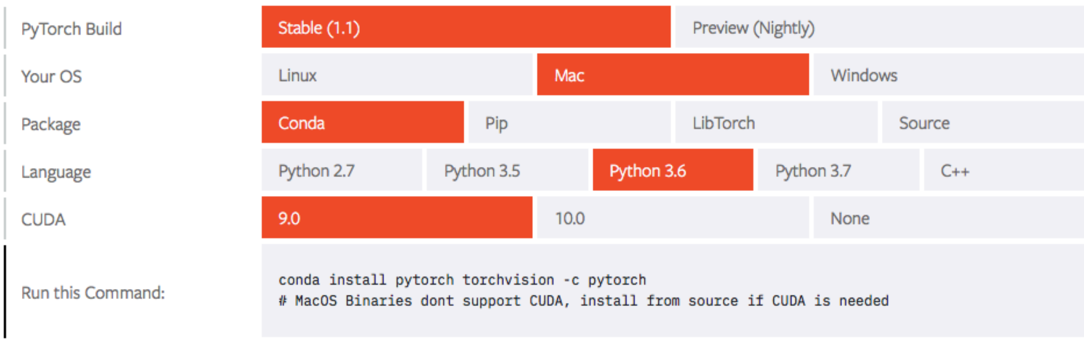


Fig 3.20 *PyTorch is used to deploy the Palces 365 CNN for Image Pre Processing*

Fig 3.21 *–basic Stack setup used for the main Scripting elements of the Workflow*

Fig 3.22 *–Mapping APIs for street level image collation, geo analysis and geo visualisation*

Fig 3.23 *–For the Cloud Infrastructure Implementation*

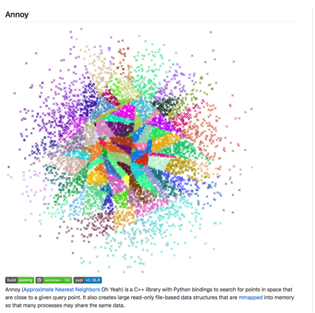


Fig 3.24 *–For the Image Similar Analysis Stage. The streamlined Annoy NN Algorhitim is created and used by Spotify.*

3.13 MobileNet and ImageNet

MobileNet V1 is a family of neural network architectures for efficient on-device image classification, originally published by Howard et al(2017). Its weights were originally obtained by training on the ILSVRC-2012-CLS dataset for image classification ("Imagenet").

ImageNet is an image dataset organized according to the WordNet hierarchy. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet, majority of them are nouns (80,000+). In ImageNet, we aim to provide on average 1000 images to illustrate each synset.

WordNet is a large lexical database of English Nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Words are grouped together based on their meanings, bar two distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

The main relation among words in WordNet is synonymy, as between the words shut and close or car and automobile. Synonyms--words that denote the same concept and are interchangeable in many contexts--are grouped into unordered sets (synsets). Each of WordNet’s 117 000 synsets is linked to other synsets by means of a small number of “conceptual relations.”

The most frequently encoded relation among synsets is the super-subordinate relation (also called hyperonymy, hyponymy or ISA relation). It links more general synsets like {furniture, piece_of_furniture} to increasingly specific ones like {bed} and {bunkbed}. Thus, WordNet states that the category furniture includes bed, which in turn includes bunkbed; conversely, concepts like bed and bunkbed make up the category furniture. All noun hierarchies ultimately go up the root node {entity}. Hyponymy relation is transitive: if an armchair is a kind of chair, and if a chair is a kind of furniture, then an armchair is a kind of furniture. WordNet distinguishes among Types (common nouns) and Instances (specific persons, countries and geographic entities). Thus, armchair is a type of chair.

The majority of the WordNet’s relations connect words from the same part of speech (POS). Thus, WordNet really consists of four sub-nets, one each for nouns, verbs, adjectives and adverbs, with few cross-POS pointers. Cross-POS relations include the “morphosemantic” links that hold among semantically similar words sharing a stem with the same meaning: observe (verb), observant (adjective) observation, observatory (nouns) .

(ImageNet Project Authors - 2018).

3.14 Scene Recognition with Places365

Understanding the world in a single glance is one of the most accomplished feats of the human brain: it takes only a few tens of milliseconds to recognize the category of an object or environment, emphasizing an important role of feed forward processing in visual recognition. One of the mechanisms subtending efficient human visual recognition is our capacity to learn and remember a diverse set of places and exemplars [11]; by sampling the world several times per second, our neural architecture constantly registers new inputs even for a very short time, reaching an exposure to millions of natural images within just a year (Zhou et Al).

Higher-level features learned by object-centric versus scene-centric CNNs are different: iconic images of objects do not contain the richness and diversity of visual information that pictures of scenes and environments provide for learning to recognize them(Zhou Et al).

There are 1.8 million train images from 365 scene categories in the Places365-Standard, which are used to train the Places365 CNNs. There are 50 images per category in the validation set and 900 images per category in the testing set.

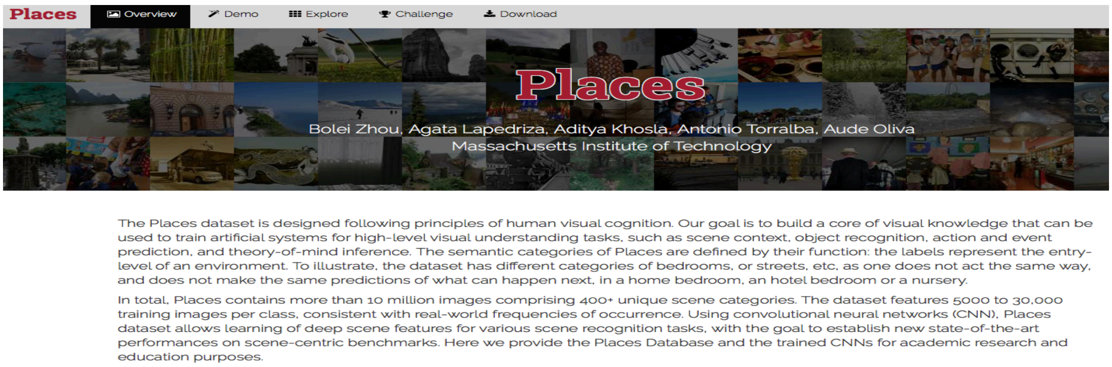


Fig 3.25 *–MIT Places pre-trained CNNs using Caffe and PyTorch.*

The Places365 Image Dataset aims to provide an exhaustive list of human and natural environment categories, bounded by spaces where a human body would fit (e.g. closet, cave). It builds on an initial list of semantic categories provided by the SUN (Scene Understanding) dataset built around a quasi-exhaustive list of scene categories with different functionalities, namely categories with unique identities in discourse .Through the use of WordNet the SUN database team selected 70,000 words and concrete terms that described scenes, places and environments that can be used to complete the phrase “I am in a place”, or “let’s go to the/a place” resulting in a corpus of 900 different scene categories after bundling together synonyms, and separating classes described by the same word but referring to different environments (e.g. inside and outside views of churches)(Zhou Et Al).

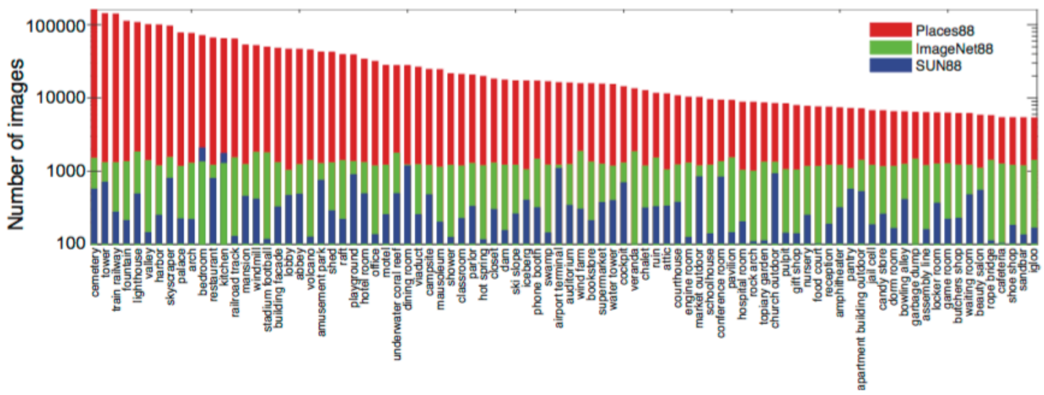


Fig 3.26 *– Comparison of the number of images per scene category for the common 88 scene categories in Places, ImageNet, and SUN datasets.*

Fig 3.27 *– Image samples from various categories of the Places Database (two samples per category).*

3.15 Visualization Software

Visualizations and Maps were produced in Python, Matplotlib, Seaborn, Tableau, Folium and ESRI ArcMap.

3.16 Local Hardware

The local implementation was deployed on an iMac (21.5-inch, Mid 2011), 2.5 GHz Intel Core i5, 16 GB 1333 MHz DDR3, AMD Radeon HD 6750M 512 MB

3.17 Scaling a Local implementation onto Google Cloud Infrastructure

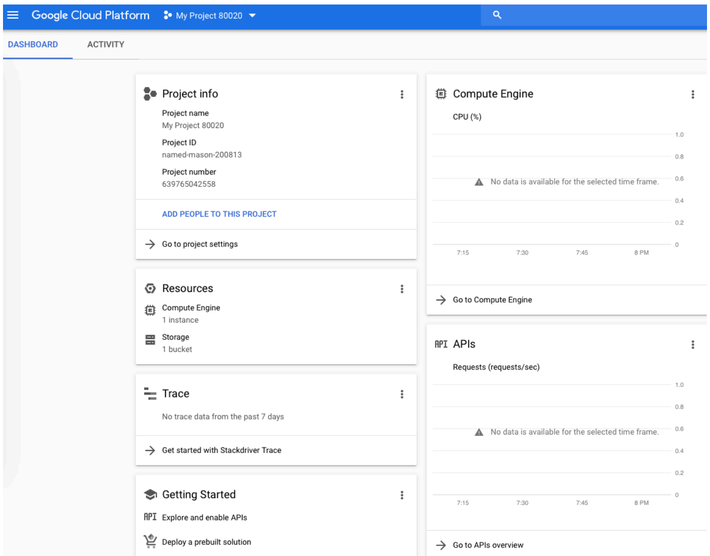
3 Deployment Scenarios were employed for this stage for performance benchmark comparison and in anticipation of a scaled solution being required to achieve the studies goals.

-Cloud GPU Instance and Data Buckets
-Cloud ML Prediction API - Cloud Dataflow and Cloud ML
-TPU

Our Image Data are hosted in Google Cloud Buckets and mounted via GCFUSE. We used an NVIDIA Tesla K80 and modified the TensorFlow scripts to utilize the TensorFlow GPU api functionality.

Google Cloud ML is the managed service for building and running machine-learning models at scale using the open source TensorFlow deep-learning framework.

TPUs are processing units that are optimized for use with the TensorFlow environment for greater speed and reliability.



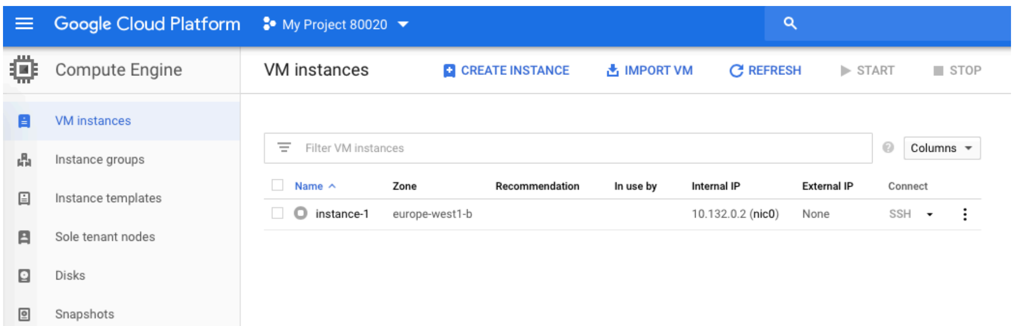


***Fig 3.28** – Our Google VM Instance and Computer Engine Console*

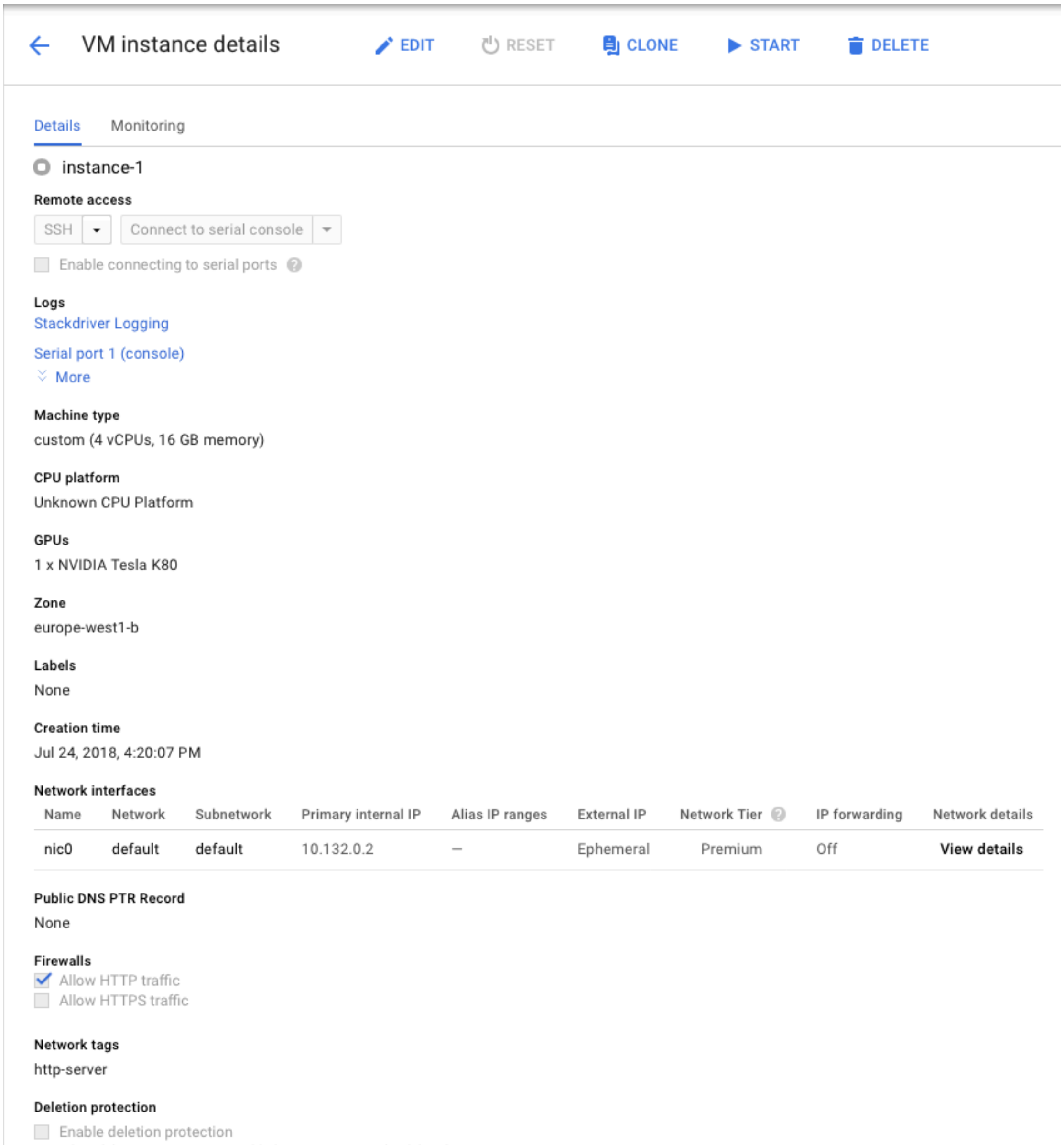


Fig 3.29 *–Spinning up a GPU Instance*

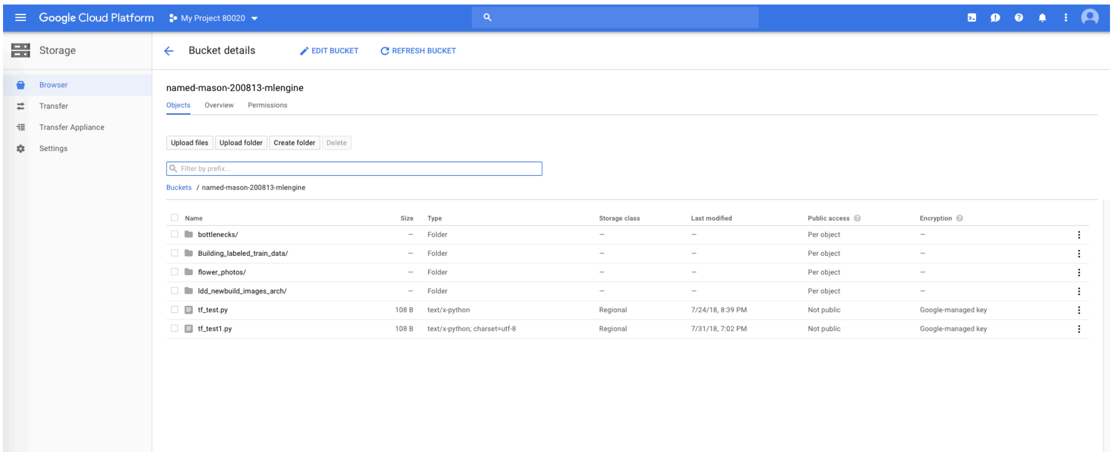


Fig 3.30 *–Our Google Cloud Bucket Repostiories*

Google Computer Engine Steps
Setup/Install
Get onto Box
Run Google Computer Client
Mount Bucket with GCFUSE
Upload Data to Bucket
Upload Scripts
Enter Parameters and Bucket Location

Fig 3.31 *– Steps for Conecting to the Bucket*

4.3.3 GPU Performance

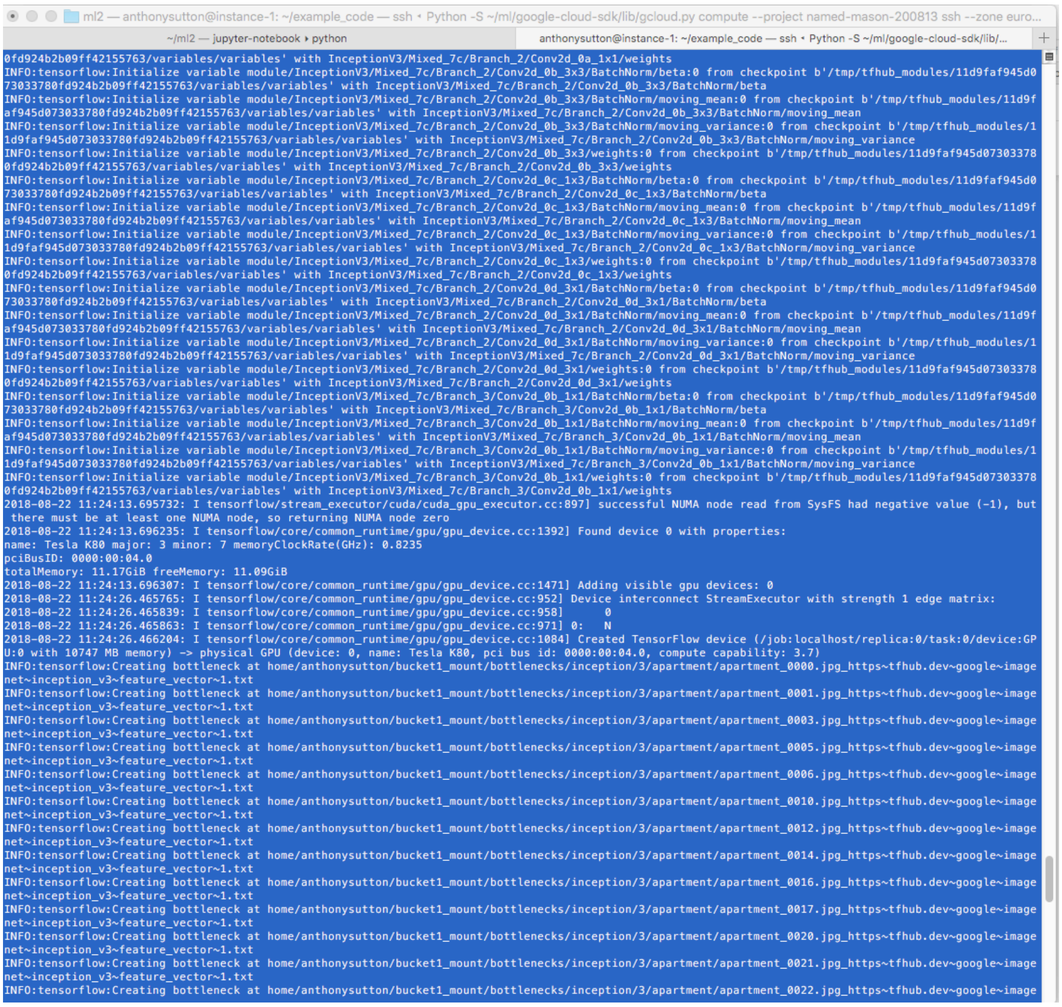


Fig 3.32 *–Using Google CME Toolkit to manage cloud deployment locally*

Running the model workflow on cloud CPU and GPU instances cloud was found to not offer great improvements in model speed and performance. This might be in part due to the number of images being processed but also to not having spend sufficient time acclimatizing to the differing requirements of the cloud based workflow solution.

For instance, accessing images stored on GCFuse accessed Mount Bucket involves accessing data which is physically or geographically far away. The propriety Google technical services on offer take this into account and appear to favor several optimized and integrated(with the google cloud architecture and infrastructure) services for deploying at scale machine learning solutions over the idea of an independent local client server system replicated wholesale onto the cloud .

The Local Machine Tensorflow and Python VM Implementation was found to be sufficient for the tasks faced with most training runs completing their step cycle within minutes if not seconds. Most of the workflow time was therefore spent working on the Locally deployed platform.

Fig 3.33 *– A small modification needs to be made to the existing tensor flow scripts to activate GPU capability.*

```python
with graph.as_default():
    (train_step, cross_entropy, bottleneck_input
     ground_truth_input, final_tensor) = add_fir
         class_count, FLAGS.final_tensor_name, b
         wants_quantization, is_training=True)

with tf.device('/gpu:0'):

  with tf.Session(graph=graph) as sess:
    # Initialize all weights: for the module t
    # and for the newly added retraining layer
    init = tf.global_variables_initializer()
```

```python
def main(_):
  # Needed to make sure the logging output is visible.
  # See https://github.com/tensorflow/tensorflow/issues/3047
  tf.logging.set_verbosity(tf.logging.INFO)

  if not FLAGS.image_dir:
    tf.logging.error('Must set flag --image_dir.')
    return -1

  # Prepare necessary directories that can be used during training
  prepare_file_system()

  # Look at the folder structure, and create lists of all the images.
  image_lists = create_image_lists(FLAGS.image_dir, FLAGS.testing_percentage,
                                   FLAGS.validation_percentage)
  class_count = len(image_lists.keys())
  if class_count == 0:
    tf.logging.error('No valid folders of images found at ' + FLAGS.image_dir)
    return -1
  if class_count == 1:
    tf.logging.error('Only one valid folder of images found at ' +
                     FLAGS.image_dir +
                     ' - multiple classes are needed for classification.')
    return -1

  # See if the command-line flags mean we're applying any distortions.
  do_distort_images = should_distort_images(
      FLAGS.flip_left_right, FLAGS.random_crop, FLAGS.random_scale,
      FLAGS.random_brightness)

  # Set up the pre-trained graph.
  module_spec = hub.load_module_spec(FLAGS.tfhub_module)
  graph, bottleneck_tensor, resized_image_tensor, wants_quantization = (
      create_module_graph(module_spec))

  # Add the new layer that we'll be training.
  with graph.as_default():
    (train_step, cross_entropy, bottleneck_input,
     ground_truth_input, final_tensor) = add_final_retrain_ops(
         class_count, FLAGS.final_tensor_name, bottleneck_tensor,
         wants_quantization, is_training=True)

  with tf.device('/gpu:0'):

    with tf.Session(graph=graph) as sess:
      # Initialize all weights: for the module to their pretrained values,
      # and for the newly added retraining layer to random initial values.
      init = tf.global_variables_initializer()
      sess.run(init)

      # Set up the image decoding sub-graph.
      jpeg_data_tensor, decoded_image_tensor = add_jpeg_decoding(module_spec)
```

4 Analysis Workflow

4.1 Workflow Overview

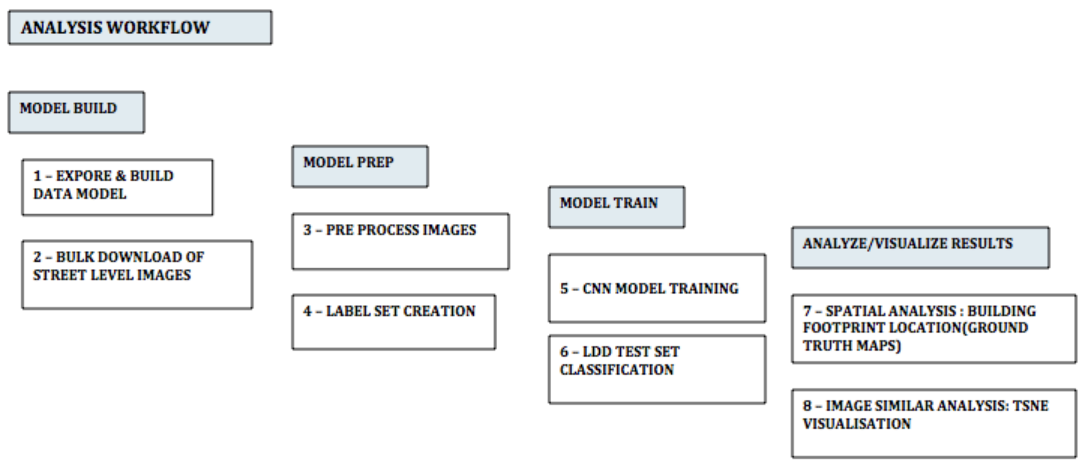


Fig 4.1 *–Workflow Stages Overview - Explore/Simplify/Iterate/Large Scale Data Model/Machine Learning Workflow*

The Overall Workflow has been designed to streamline and manage the requisite steps of an Urban Image Classification Task whilst also providing a reusable set of tools for more generic CNN Tasks using the Tensorflow API and platform in general.

The basic goal of the Image Classification Task is to train a Neural Network to correctly identify and describe Images. The Secondary goal of our study is to extend this classification process into Image Analysis of the model weights trained on our London Buildings dataset.

An important part of any Machine Learning exercise and for an Image Classification in particular, is benchmarking. This translates into the Metric Values of Accuracy and Speed. The workflow has been implemented to provide the means to monitor the component tasks (e.g. image pre processing, label creation, CNN training), to measure metrics from different variable or parameter setting training runs and ultimately the metric benchmark performances for comparison with other CNN Classification Tasks and studies from the wider Research field.

A Jupyter Python 3 Script was created to accompany each stage. The steps covered in each stage and functionality provided are outlined in the workflow diagrams below along with descriptions of the workflow tools and utilities created for each stage.

4.2 Bespoke Workflow Tools

Maintaining an ongoing and rapid “Ground Truth” referral process played a large role in the overall workflow and the accompanying software tools for the project (e.g. the spatial locator and building inspector, image thumb gallery inspector, sample selector widgets and quick spin function tools) were designed with this in mind. This would ease the often-time-consuming process of matching large amounts of images to a verified geo reference, validating addresses(we needed to be sure an address record and its linked features in the data model were factually correct throughout the image classification verification stage), checking building development lifecycle data and in troubleshooting and assessing different label set creation approaches as well as in examining CNN training accuracy and ensuring that the machine learning components were functioning correctly throughout the workflow.

At each stage of the candidate label process sample size charts were used for a quick fire inspection assessment of the candidate data. To assist in this process a sample size dashboard was embedded into each section of the Jupyter Script. In Addition several other tools were created and employed throughout to manage and visualize throughput and output and which eased the numerous repeated runs through the workflow with differing parameters or variation. Please explore the Juypter Scripts for more detail.

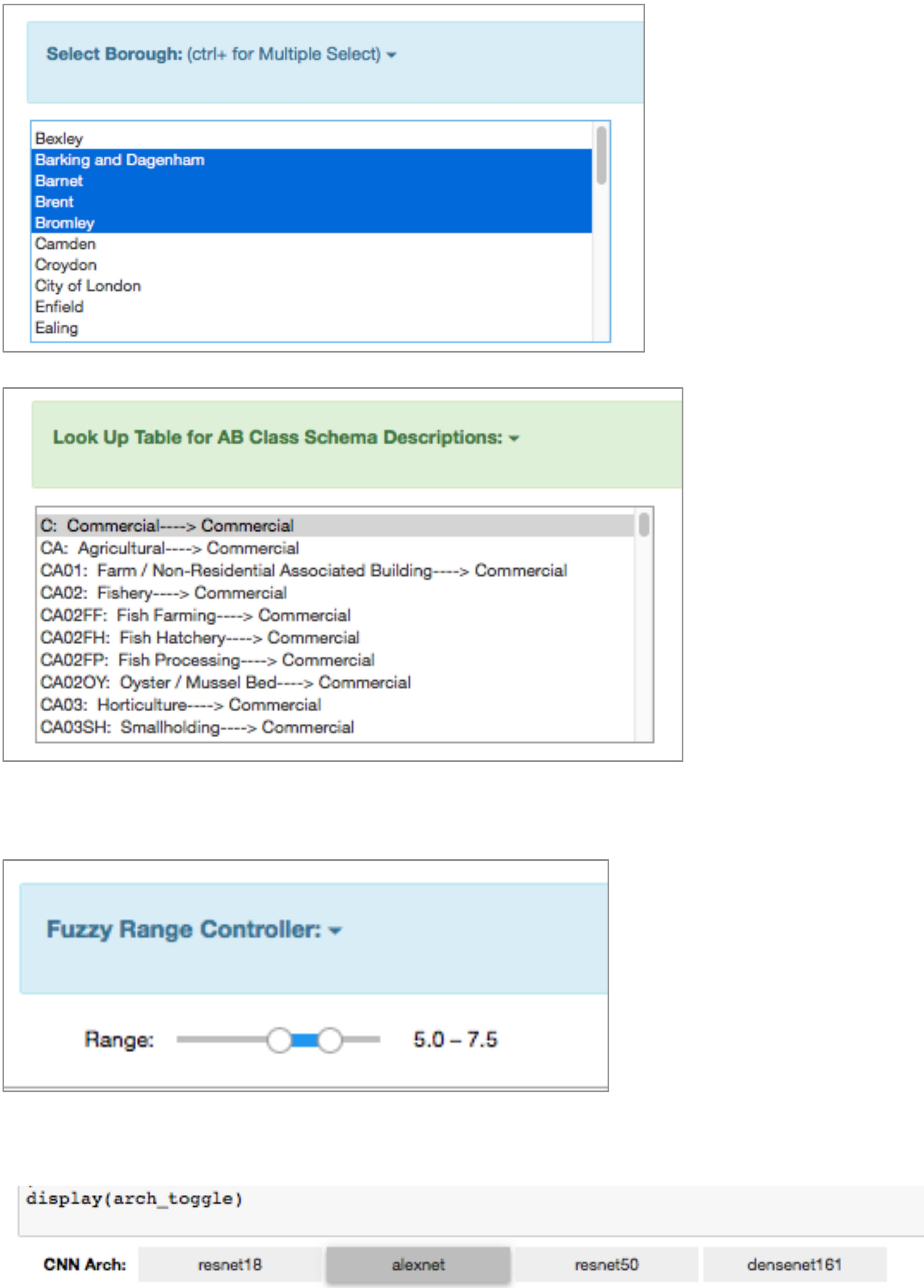


Fig 4.2 - *Python Jupyter Widgets for Handling User Input*

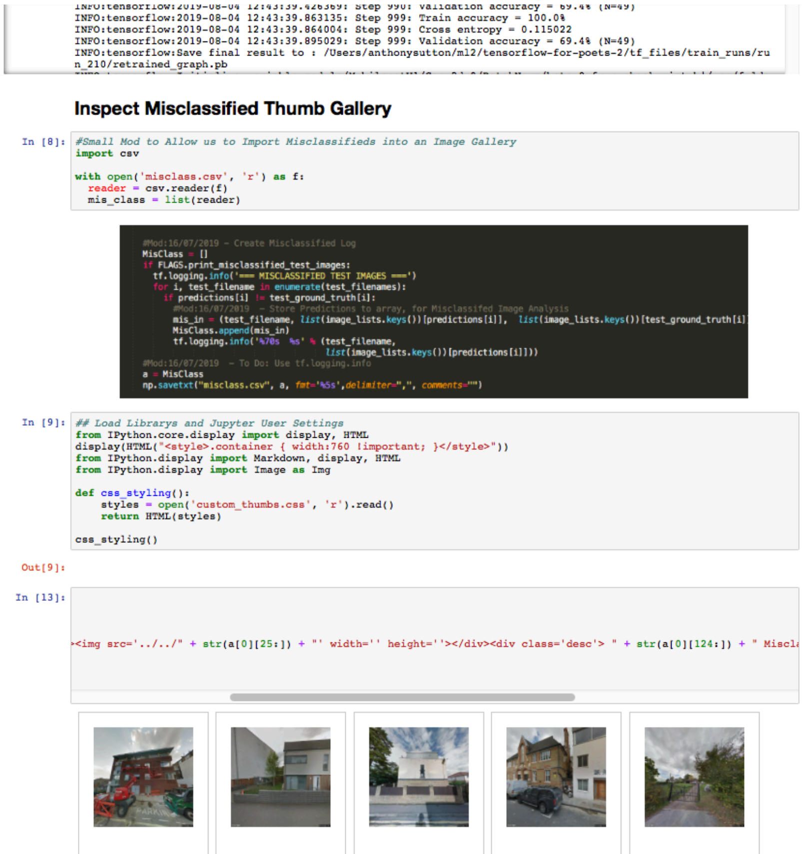
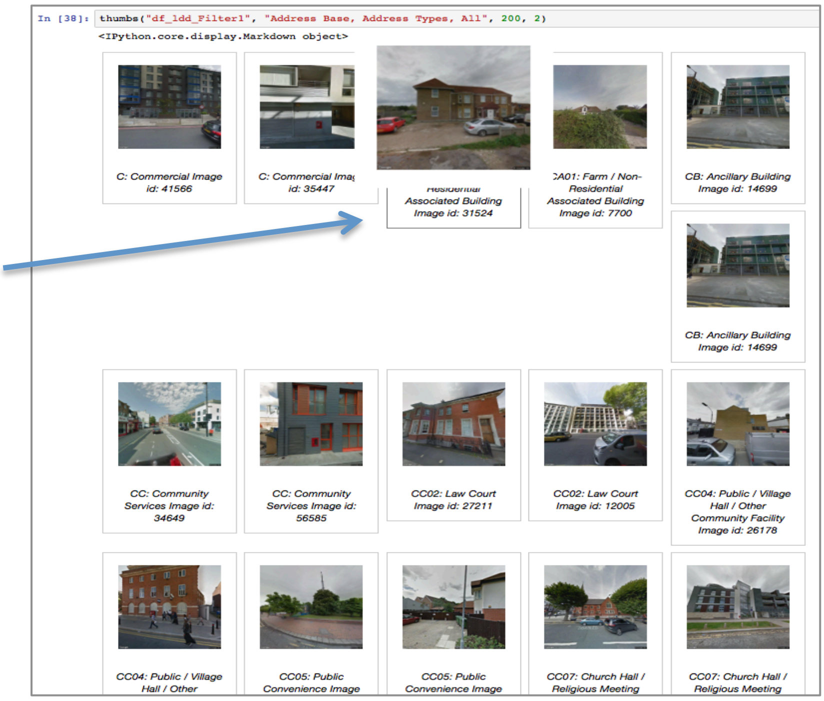
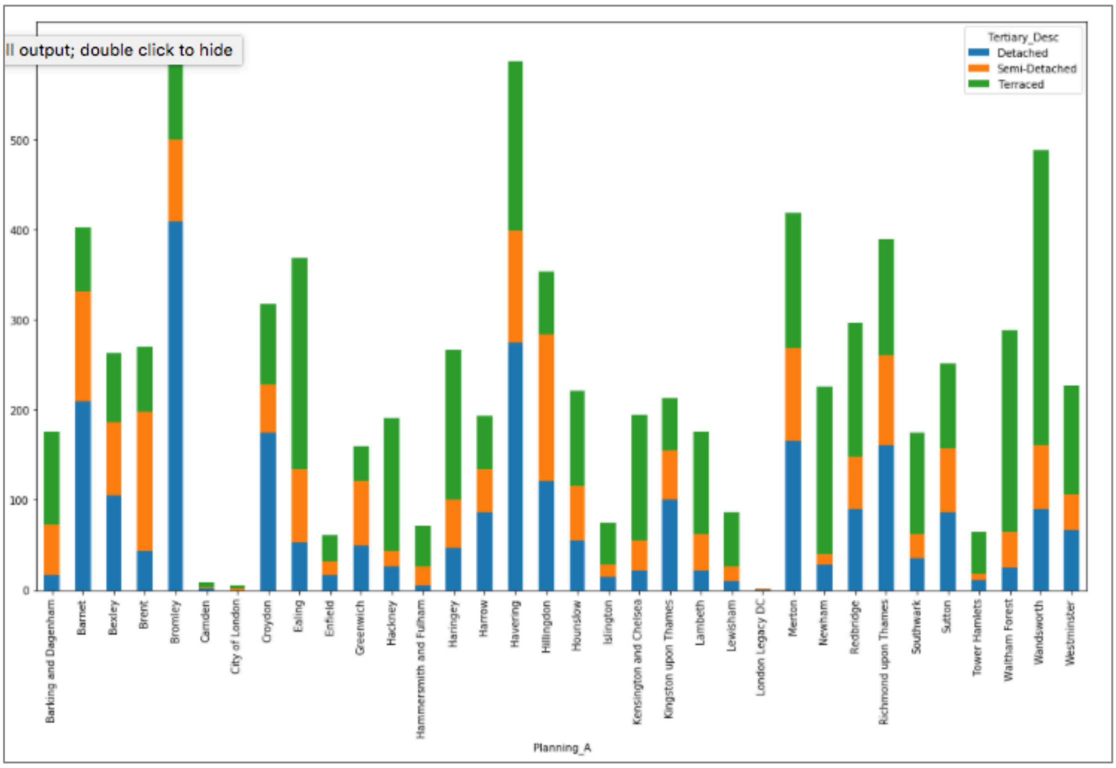


Fig 4.3 *– Hover Magnifier **Thumb Gallery** for managing and inspecting large amounts and sets of images and building types in the data validation stages.*





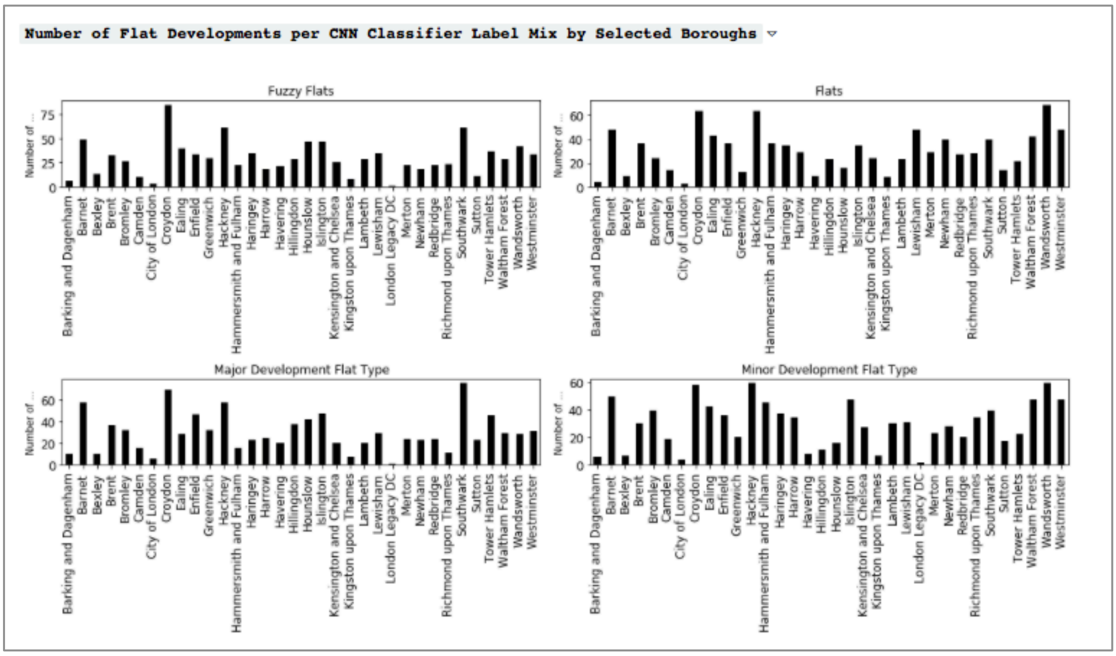


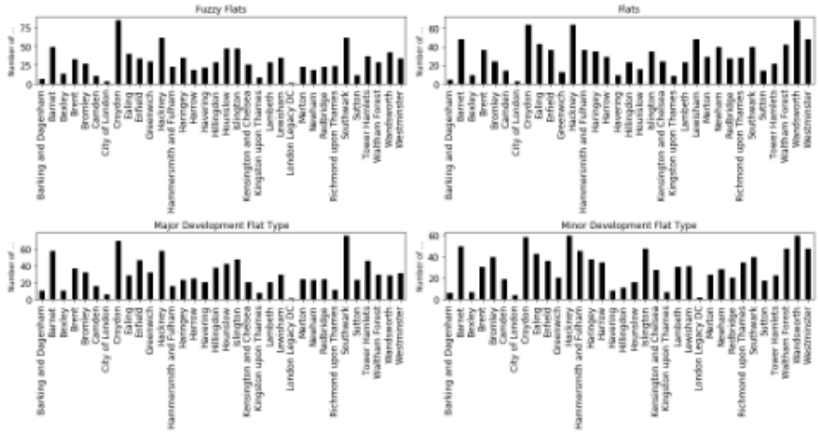
Fig 4.4 – *Label Set Creation* ***Dashboards*** *for Category Sample Spread Checking during the Label Creation Process*

```
msk = np.random.rand(len(df_cat_flats_minor_200)) < 0.8
df_cat_flats_minor_200_train = df_cat_flats_minor_200[msk]
df_cat_flats_minor_200_test = df_cat_flats_minor_200[~msk]

#Fuzzy Dev Flat Type
msk = np.random.rand(len(df_cat_flats_fuzzy_units_200)) < 0.9
df_cat_flats_fuzzy_units_200_train = df_cat_flats_fuzzy_units_200[msk]
df_cat_flats_fuzzy_units_200_test = df_cat_flats_fuzzy_units_200[~msk]

msk = np.random.rand(len(df_cat_flats_fuzzy_units_1000)) < 0.9
df_cat_flats_fuzzy_units_1000_train = df_cat_flats_fuzzy_units_1000[msk]
df_cat_flats_fuzzy_units_1000_test = df_cat_flats_fuzzy_units_1000[~msk]
```

In [101]:

```
printmd('<br><br><b>`Number of Flat Developments per CNN Classifier Label Mix by Selected
fig=plt.figure(figsize=(15,8))
plotdat_group(df_cat_flats_fuzzy_units_1000,'Planning_A', 221, 'Fuzzy Flats')
plotdat_group(df_cat_flats_1000,'Planning_A', 222, 'Flats')
plotdat_group(df_cat_flats_major_1000,'Planning_A', 223, 'Major Development Flat Type')
plotdat_group(df_cat_flats_minor_1000,'Planning_A', 224, 'Minor Development Flat Type')
plt.tight_layout()
plt.show()
```

```
<IPython.core.display.Markdown object>
```



As can be seen from the above, we now have 3 Variations of Flat Class Definition with which to fine tune and determine the greatest recall accuracy with the Inception CNN architecture Classfier.

The latter 2 categories focus on scale of development, as outlined, for example, in The Town and Country Planning (Development Procedure) (England) Order 2010 Management. Criteria for Developments to be considered in the Major Category include *10+ dwellings / over half a hectare / building(s) exceeds 1000m²*. Minor Dwellings are therefore *1-9 dwellings (unless floorspace exceeds 1000m² / under half a hectare*. Ultimately we are seeking to differentitate between Large Blocks of Flats and smaller Building development Object Types.

Note that we have no properties in the Flat_Conversion Category, which accords with our study's focus on New Build Development. These have been excluded as part of the Data Model Build Stage. Please note that our rather clumsy pattern matching may inadvertandly also be excluding conversions that are relate not just to single flat, e.g. conversion of offices B1a to C3 residential use(See the Town and Country Planning (General Permitted Development) (England) Order 2015)

Our Set Controls in this category are provided by text from the proposal field and Residential Unit Counts, both provided by the LDD

Inspect Image Data:

In [ ]:

```
thumbs("df_cat_flats_major_1000", "Major Development", 100, 2)
thumbs("df_cat_flats_minor_1000", "Minor Development", 100, 2)
thumbs("df_cat_flats_fuzzy_units_1000", "Fuzzy Set Control", 100, 2)
thumbs("df_cat_flats_no_conv", "No Conversion", 100, 2)
```

2.6 - Create Office Property Type Label:

Fig 4.5 *– Label Set Creation Dashboard(as seen from within the Jupyter Script,* ***Flat*** *Buiding Type Label Creation Cells)*

In [121]:

```
printmd('<br><br><b>`Number of New Build By Property Type in ' + str(Borough_Selection) +
df_proptype = df_cat_commercial_retail.groupby(['Planning_A', 'ClassificationCode'])['Plan
df_proptype.plot(kind='bar', stacked=True, figsize=(18, 10));

printmd('<br><br><b>`Number of Retail Developments per CNN Classifier Label Mix by Selecte
fig=plt.figure(figsize=(15,12))
plotdat_group_null(df_cat_commercial_retail_1000_train,'Planning_A', 221, 'Retail')
plotdat_group_null(df_cat_commercial_retail_1000_train,'Planning_A', 222, 'Retail - London
plotdat_group_null(df_cat_commercial_retail_1000_major_train,'Planning_A', 223, 'Retail De
plotdat_group_null(df_cat_commercial_retail_1000_minor_train,'Planning_A', 224, 'Retail De
plt.tight_layout()
plt.show()
```

```
<IPython.core.display.Markdown object>
```

Out[121]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x11787b320>

<IPython.core.display.Markdown object>
```

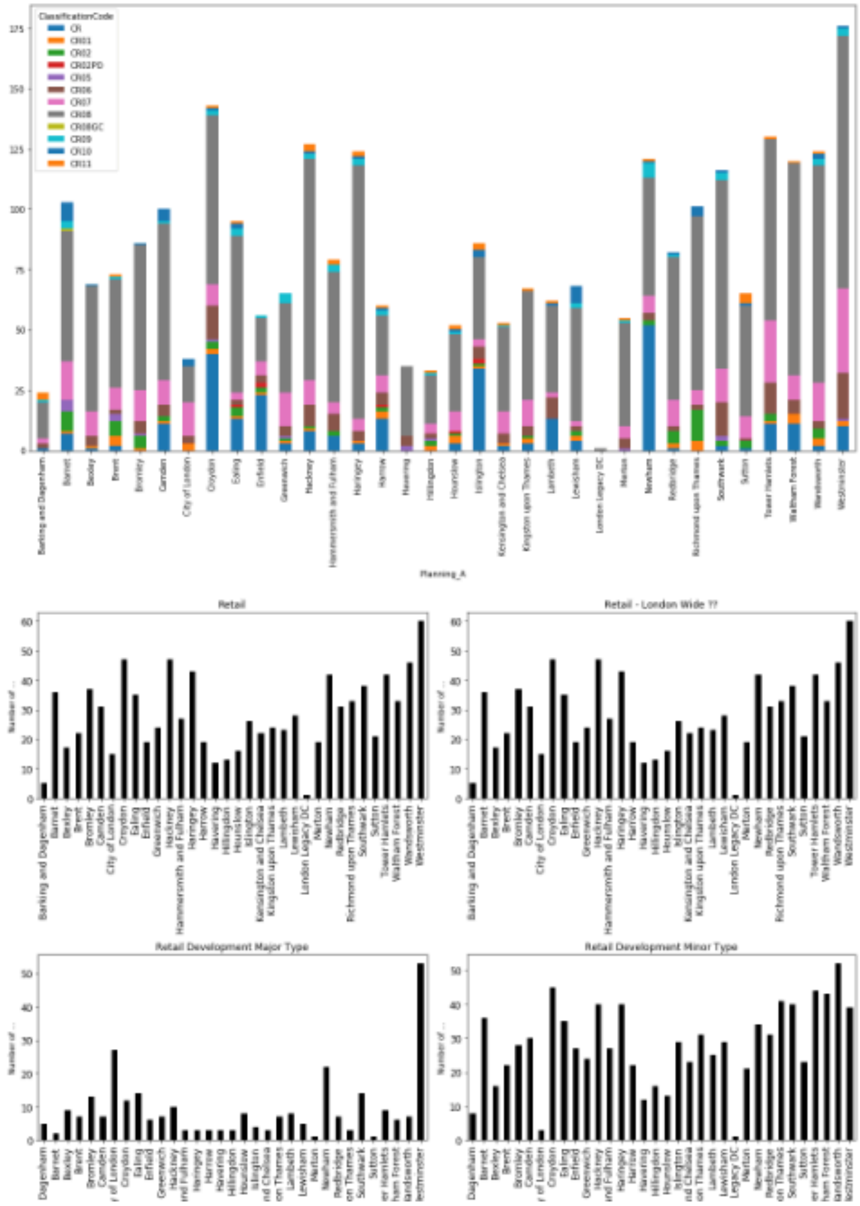


__Fig 4.6.__ – Label Set Creation Dashboard(as seen from within the Jupyter Script, __Retail__ Buiding Type Label Creation Cells)

4.2 Workflow Stages 1 through to 8 - Fig 4.6 provides a visual overview of each workflow stage. More detailed exposition of the processes and content of each workflow stage can be found in the sections that follow and in the annotated Code/Jupyter Scripts archive located in the GitHub appendices.