

Article

Grammar Guided Genetic Programming for Network Architecture Search and Road Detection on Aerial Orthophotography

Víctor de la Fuente Castillo ^{1,*}, Alberto Díaz-Álvarez ^{2,†}, Miguel-Ángel Manso-Callejo ^{3,†}
and Francisco Serradilla García ^{2,†}

¹ Centro de Informática y Comunicaciones, E.T.S.I. de Sistemas Informáticos, Universidad Politécnica de Madrid, 28031 Madrid, Spain

² Departamento de Inteligencia Artificial, E.T.S.I. de Sistemas Informáticos, Universidad Politécnica de Madrid, 28031 Madrid, Spain; alberto.diaz@upm.es (A.D.-A.); francisco.serradilla@upm.es (F.S.G.)

³ Departamento de Ingeniería Topográfica y Cartografía, E.T.S.I. Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, 28031 Madrid, Spain; m.manso@upm.es

* Correspondence: victor.fuente@upm.es

† Current address: Universidad Politécnica de Madrid (Campus Sur), Calle Alan Turing s/n (Ctra. de Valencia, Km. 7), C.P. 28031 Madrid, Spain.

Received: 1 May 2020; Accepted: 3 June 2020; Published: 6 June 2020



Featured Application: This system can be applied to automatically design the architecture of deep artificial neural networks for any given labeled dataset.

Abstract: Photogrammetry involves aerial photography of the Earth's surface and subsequently processing the images to provide a more accurate depiction of the area (Orthophotography). It is used by the Spanish Instituto Geográfico Nacional to update road cartography but requires a significant amount of manual labor due to the need to perform visual inspection of all tiled images. Deep learning techniques (artificial neural networks with more than one hidden layer) can perform road detection but it is still unclear how to find the optimal network architecture. Our main goal is the automatic design of deep neural network architectures with grammar-guided genetic programming. In this kind of evolutive algorithm, all the population individuals (here candidate network architectures) are constrained to rules specified by a grammar that defines valid and useful structural patterns to guide the search process. Grammar used includes well-known complex structures (e.g., Inception-like modules) combined with a custom designed mutation operator (dynamically links the mutation probability to structural diversity). Pilot results show that the system is able to design models for road detection that obtain test accuracies similar to that reached by state-of-the-art models when evaluated over a dataset from the Spanish National Aerial Orthophotography Plan.

Keywords: grammar evolution; deep learning; network architecture search; Grammar-guided genetic programming

1. Introduction

Two of the main problems with remote sensing information as noted by Risojević et al. [1] are that the high volume of data exceeds by far the capabilities of human analysis (by manual revision) and at the same time is usually crucial to perform classification of said data. Under this category of data, large-scale aerial image processed as orthophotography is a useful source of information for many domains. To give some examples of the broad array of applications we can find, there have been systems developed for the detection of coastline changes [2], snow avalanches [3], fires [4],

bodies in disaster sites [5], trees [6], seedlings [7], roofs [8], transmission towers [9,10], vehicles [11,12], photovoltaic arrays [13,14], vegetation and buildings [15]. We focus on road detection on aerial images being it an important subject, among other things, due to the need to constantly update road maps. This importance is clear by the volume of research conducted and applied systems developed on this domain.

As an object of detection, Fortier et al. [16] divided the road characteristics into four main groups: spectral properties (those regarding surface characteristics), geometric (such as width and curvature), topological properties (those derived from their role as part of a transportation network), and contextual (e.g., different types of roads may have different restrictions to their shape, size or materials). Put in the context of aerial image we can find many complications related to those groups or properties some even caused by the geographic context of the roads. One of these problems as shown in Figure 1a,b is due to surrounding terrain having spectral properties similar to the secondary roads shown. Other common problem are occlusions due to their contextual properties (e.g., secondary mountain roads). Here, we show an example of said occlusion at forest areas (Figure 1c–f), on the first case due to the shadows and the rest the trees themselves covering the road. Other problems such as some natural or artificial structures being similar to roads are shown in Figure 1g–o.

Some road detection systems rely entirely on non-neural models: Wang et al. [17] combined texture information extraction with a knowledge-based road extraction and a post-processing step for noise elimination; Mattyus et al. [18] applied Structural Support-Vector Machines (S-SVMs); and Tsochantaridis et al. [19] and Zhou et al. [20] used GraphCut [21]. Although we can find many other systems that rely on road characteristics analysis [16–18,20,22–28], most recent research has favored the application of deep artificial neural networks (ANNs). We find some systems include ANNs as part of their detection pipeline (e.g., [29–33]) or just preprocess the input images before they are feed to the network as in [34]. It is frequent to rely on a deep ANN to perform the whole road detection/segmentation process by either developing a novel network or using a preexisting architecture. Some researchers use custom network architectures [35,36], others customize well-known models [30,37–40], use those well-known models with transfer learning [41,42] or combine them into road-detection ensembles [43].

Systems that apply deep learning techniques to detection on such images can use quite different network dimensions and architectures. We can find ones that go from relatively small convolutional networks, such as the one in [44], to others using models such as the Inception-V3 architecture [45] combined with transfer learning, as in [14] for solar panel detection. To try and solve the need to find a quasi-optimal ANN architecture automatically, Network Architecture Search (NAS) has been extensively researched by using many different approaches but has proven to be a complex task given the dimensions of the search space. Some examples of such approaches include the use of reinforcement learning techniques, as seen in [46–48], or the application of evolutive algorithms (e.g., [49–60]).

Some systems use grammars to define the rules to obtain valid structures that can encode either the neuron connectivity matrix and some other form of network topology [54,56,57] or other higher level expression of the operations being performed by the network layers and their connectivity [52,53,55,58–61]. Some of those systems (e.g., [52,53]) work by generating architecture expressions with high-level operations (such as convolutions, activation functions, dropout, etc.). In [53], the architecture is sequential in term of layers, while Assunção et al. [52] allowed more complex connectivities (parallel branches inside the network).

The present work is conducted inside the Cartobot project for the Instituto Geográfico Nacional (IGN, Spain) to perform automatic design of deep neural networks for road detection on aerial images, guiding the search with information about useful network structural patterns. This project is of real application on a dataset constructed (by IGN and Cartobot personnel) from Spanish National Plan for Aerial Orthophotometry (PNOA) aerial orthoimages. The two main research goals are: (a) design a grammar that can successfully encode the most typical high-level architectural blocks of deep ANNs; and (b) build a NAS system capable of designing deep ANNs using that grammar for the purpose

of successfully detecting roads on aerial images. We divert from the approach of Assunção et al. [52] in that our grammar is going to define specific high-level blocks, which is discussed below, instead of allowing fully free connectivity for the generation of the blocks with parallel branches. We also work on a dynamic mutation operator linked to the structural diversity of solutions (population) being contemplated at each instant.

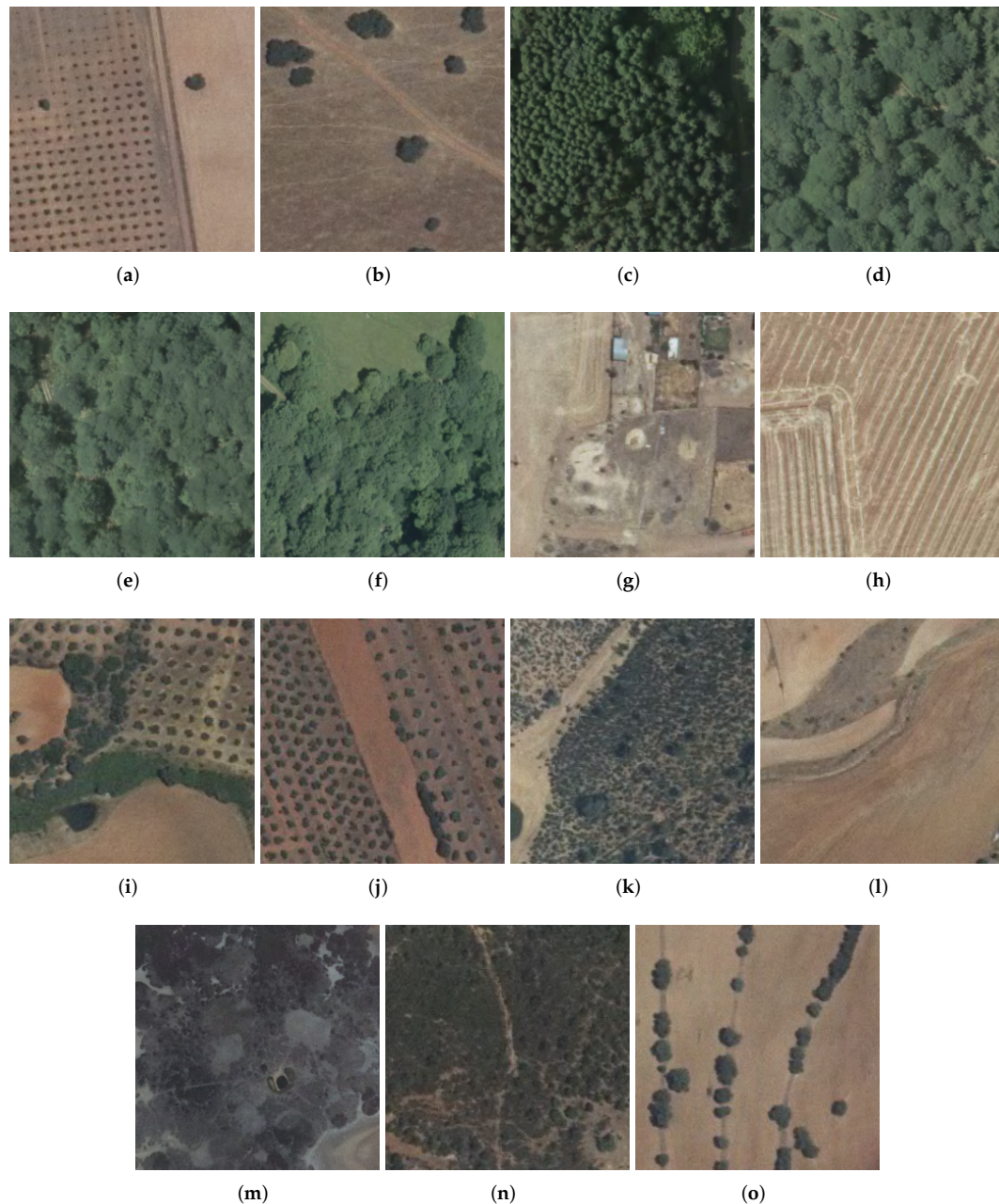


Figure 1. Examples of potential issues: (a–f) roads with complex contextual terrain or partial occlusions; and (g–o) images not containing roads that could be hard to classify.

We start by reviewing some of the well-known network models (AlexNet [62], VGG [63], GoogleLeNet [64], ResNet [65], Inception-v3 [45], Inception-v4 and Inception-ResNet [66], DPN [67] and ResNeXt [68]) used by top scoring results on the Imagenet Large Scale Visualization Recognition Challenge (ILSVRC) [69] (editions from 2012 to 2017) where many different architectural patterns can be found. As mentioned, by defining a meta-grammar able to condense some of such structural rules,

we aim to guide an evolutive NAS system to find networks capable of solving the road detection problem on aerial orthoimages. The problem of road detection (including secondary roads) on such orthoimages has, as explained above, the added difficulty of natural or artificial structures that can be mistaken with roads or transportation infrastructures. Some samples of those formations can be seen in Figure 1g–o with various levels of complexity. On the other hand, many different types of roads are being considered and are required to be detected by the system. We have a set of sample orthoimages with their corresponding numerical topographical cartography at a scale of 1:25,000 (BTN25) to allow for easier identification of the roads by the reader (see Figure 2).

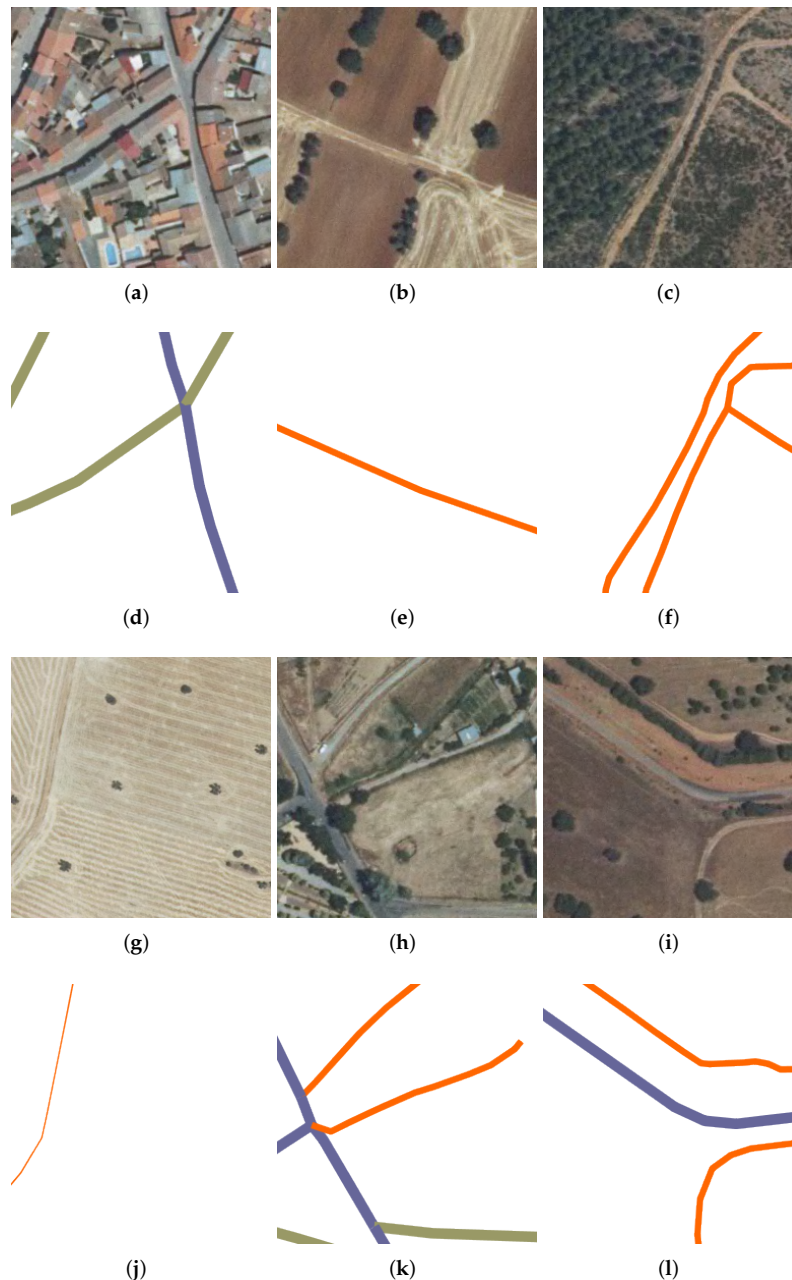


Figure 2. Sample images containing roads (a–c,g–i) and their corresponding road cartography (d–f,j–l).

2. Methods

Our system, VGNet, uses a Grammar Guided Genetic Programming (GGGP) [70] to solve the NAS problem. GGGP is a family of evolutionary algorithms that encode solutions to a given problem

as tree structures (constrained by a set of structural rules specified with a grammar) and follows the same basic process of a genetic algorithm with some variations. The reasons for selecting this approach are: evolutionary algorithms easily allow for a high-level of parallelism [55] and the use of a grammar avoids exploring invalid solutions and allows for domain knowledge to be included to guide the search. This kind of systems can find quasi-optimal solutions for a given problem when finding the global optimum is not possible due to computational complexity.

The general GGGP evolutive process follows these steps:

1. We define a coding system to represent the candidate solutions (individuals) for a given problem and a context-free grammar that defines restrictions to their structure.
2. A population (a set) of initial solutions is created following the grammar rules.
3. We check the fitness (a measure of how well they solve the problem at hand) of the individuals to see if we have found an acceptable solution (given some stop criterion or goal to reach) and should stop the process or continue the search.
4. If the stop criterion is not met, we create a new population as follows:
 - (a) Select solutions (called parents), usually by pairs, with regard to their fitness value. Better solutions have higher chances of being selected.
 - (b) With certain probability, combine the parents (or leave them unchanged) to obtain new solutions (offspring) using a crossover operator. The goal here is to combine the information contained in each parent to try and find better solutions. In our scenario, subtrees are exchanged between the parents to generate the offspring.
 - (c) With certain probability, the offspring individuals are checked to add some random variations (constrained once again by the grammar rules) to improve the exploration of new solutions (new areas of the search space).
5. We go back to step 3 with the new population we created.

For VGNet, the individuals (candidate solutions) represent an ANN constrained by a context-free grammar of highly used network architectural building blocks. The system has the internal workflow shown in Figure 3. Each individual generated is a derivation tree and represents a neural network architecture codified using an expression language designed to be easily read. We refer to ANNs with more than one hidden layer as deep learning models, or deep neural networks.

The operators used during the evolution process are:

- Individual generation: Based on Grammar-Based Initialization Method (GBIM) [71]. GROW [72] was used initially but was later discarded in favor of GBIM.
- Mutation: Based on Grammar-Based Mutation (GBM) [73] and a variation of it of our own design, Schema Diversity Factor (SDF) mutation.
- Probabilistic Tournament Selection [74] using the formula from [75].
- Crossover: Whingham (WX) [70].
- Generational population replacement.

Methods used by initialization and mutation are based around GBIM and GBM with the following modifications. We removed for all non-terminals the restriction of having at least one terminal node only production (hereafter, we refer to those as default productions). This affects the guaranteed uniformity of depth level during random generation studied by the authors of the original method [71]. In addition, when creating a derivation tree, a random depth is chosen from the interval on the interval $[min_depth, max_depth]$ (original GBIM uses the interval $[1, max_depth]$). These modifications allow for an easier to read meta-grammar. In addition, although the formal study is outside the scope of our present work, we believe that the use of the cited default productions on the meta-grammar can lead to the introduction of a bias in their favor during the derivation tree generation.



Figure 3. VGNet: network architecture design workflow.

The following sections detail the codification scheme (Section 2.1), with the detailed meta-grammar in Section 2.1.2. The SDF mutation operator we apply is explained in Sections 2.2 and 2.3 and, finally, the fitness in Section 2.4.

2.1. Codification Scheme

2.1.1. Expression Language

To encode the network structures, we define an expression language capable of simply representing many of the classification architectures found in the state of the art with some exceptions (e.g., DenseNet [76] is not supported due to the syntax we use for parallel processing branches). At this point, we only consider architectures with one input point and one classification output (architectures such as Faster-RCNN [77] with multiple outputs at different levels of the network are not supported in the current version of the system).

The expressions have the following tokens as basic components:

- “in”, “out”: They represent the network input and output, respectively.
- “[”, “]”: Start and end of parallel branches of processing that share the same input (branches are separated by “,”).
- “direct”: Connection between two network points without any operator being applied (residual-like connection).
- Processing nodes with two main types:
 - Traditional operators: convolutions (“conv”), max or average poolings (“mp” or “ap”), “dropout” [78], fully-connected layers (“fc”), “softmax” and “flatten”.
 - Aggregation operators: concatenation (“concat”) or zero-padded sum (“sum”). The last one adds the features of all branches with the same channel index and concatenates the rest.

Both aggregation operators are applied immediately after the closing of parallel processing branches to unify the results.

Some of the traditional operators use a parametric syntax with format **operator[-parameter][0,n]**. The specific number of required parameters varies between operators:

- “conv-i-j-k l”: i (dimension), j (number of generated features), k (stride) and l (normalization used). Currently, only Batch Normalization (“bn”) [79] or no normalization are supported by our system. We also performed some tests including both regular (“conv”) and separable convolutions (“sepconv”).
- “mp-m-k” and “ap-m-k”: m (dimension) and k (stride).
- “dropout-r”: r (rate).
- “fc-s”: s (size).

Some important notes: “flatten” is always present before the first FC layer and all layers use leaky-ReLU [80] activation (with the only exception of the output layer that uses softmax). Regarding the “in”, “softmax” and “out” nodes, the input and output dimensions are not specified on the operators themselves since they are problem dependent and configured at library level.

2.1.2. Meta-grammar

To guide the evolutionary search, the allowed network architecture is constrained to:

1. Input layer: Dimensions are given by the images of the dataset being used.
2. Sequence of high-level structural blocks. Each block is chosen from the following set:
 - Convolutional block. Sequence of convolution, optional batch normalization, dropout and optional pooling.
 - Inception-based blocks. Parallel convolutional branches with their outputs being concatenated (all branches are constrained to apply the same spatial dimensional reduction or none).
 - ResNet-based blocks. Consists of two parallel processing branches. A convolutional sequence and residual connection with their outputs being aggregated by zero padded sum (to allow for different number of features between the two branches).
 - Inception-ResNet-based blocks. Here, we have an Inception-based block and a residual connection with aggregation of both their outputs via a zero padded sum.
3. Sequence of fully connected layers (leaky-ReLU activation function) alternated with dropouts. Before the first FC layer, a flatten operation is automatically performed.
4. Output layer: softmax layer with one output per class.

It is worth mentioning that the dropout operator presence is forced allowing a value of $rate = 0$ to account for it being inactive. That way the operator block is always present and can be easily turned on/off via mutation or crossover. Dimensions higher than three are not used for the convolutions and poolings in light of the results obtained by Szegedy et al. [45] using factorized convolutions to replace those with higher dimensions. Limit for number of FC layers exists mainly due to their computational cost compared to the convolutional ones.

At a lower level, the meta-grammar has the following parameters:

- FC layers; maximum number of layers and set of dimensions allowed: **fM**, **fS**.
- Convolutions; sets of dimensions and number of features allowed: **cD**, **cF**.
- Pooling; sets of pooling types and dimensions: **pT** (e.g., MP and AP), **pD**.
- Convolution and pooling; set of valid strides: **cpS**.
- Normalizations; set of allowed normalizations: **norm**.

- Dropout allowed rates: **dR**.

The meta-grammar is expressed as a context-free grammar described here in Backus–Naur Form (BNF). To reduce the text extension, we use the contraptions Block (B), Network (N, NS for plural), Parallel (PRL), and Reduce Dimension (RD):

$$G_{fM,fS,cD,cF,cpS,pT,pD,norm,dR} = \{N_{cpS}, T_{fS,cD,cF,cpS,pT,pD,norm,dR}, P, S=ANN\}$$

$$\begin{aligned} N_{cpS} = \\ \{ ANN, HIGHLVL_N, BLOCKS, RD, FC, FC_B, RESN_B, \\ INCEPT_B, INCEPT_RESN_B, CONV_N, CONV_B, \\ CONV_N_S1, CONV_B_S1, PRL_CONV_NS, PRL_B_RD, \\ PRL_CONV_NS_RD, RD, CONV, POOL, DROPOUT \} \\ \cup \{ CONV_Sk, POOL_Sk \} \forall k \in cpS \end{aligned}$$

$$\begin{aligned} T_{fS,cD,cF,cpS,pT,pD,norm,dR} = \\ \{ 'in', 'out', 'flatten', 'softmax', 'direct', 'concat', 'sum', 'dropout', '[', ']', ',', 'direct' \} \\ \cup \{ 'conv-i-j-k-l' \} \forall i \in cD, \forall j \in cF, \forall k \in cpS, \\ \forall l \in norm \\ \cup \{ 'p-m-k' \} \forall p \in pT, \forall m \in pD, \forall k \in cpS \\ \cup \{ 'dropout-r' \} \forall r \in dR \\ \cup \{ 'fc-s' \} \forall s \in fS \end{aligned}$$

Production rules P are the following:

$$\begin{aligned} \langle ANN \rangle &::= 'in' \langle HIGHLVL_N \rangle \langle RD \rangle 'flatten' \langle FC \rangle 'softmax' 'out' \\ \langle HIGHLVL_N \rangle &::= \langle BLOCKS \rangle \\ &\quad | \langle HIGHLVL_N \rangle \langle HIGHLVL_N \rangle \\ \langle BLOCKS \rangle &::= \langle CONV_B \rangle \\ &\quad | \langle RESN_B \rangle \\ &\quad | \langle INCEPT_B \rangle \\ &\quad | \langle INCEPT_RESN_B \rangle \\ \langle FC \rangle &::= \langle DROPOUT \rangle \langle FC_B \rangle \langle DROPOUT \rangle \\ &\quad | \dots \\ &\quad | \langle DROPOUT \rangle \langle FC_B \rangle_0 \langle DROPOUT \rangle_0 \dots \langle FC_B \rangle_{fM} \langle DROPOUT \rangle_{fM} \\ \langle CONV_N \rangle &::= \langle CONV_B \rangle | \langle CONV_B \rangle \langle CONV_N \rangle \\ \langle CONV_B \rangle &::= \langle CONV \rangle \langle DROPOUT \rangle \\ &\quad | \langle CONV \rangle \langle DROPOUT \rangle \langle POOL \rangle \\ \langle CONV_B_S1 \rangle &::= \langle CONV_S1 \rangle \langle DROPOUT \rangle \\ &\quad | \langle CONV_S1 \rangle \langle DROPOUT \rangle \langle POOL_S1 \rangle \\ \langle CONV_N_S1 \rangle &::= \langle CONV_B_S1 \rangle \\ &\quad | \langle CONV_B_S1 \rangle \langle CONV_N_S1 \rangle \\ \langle RESN_B \rangle &::= '[' \langle CONV_N_S1 \rangle ',' 'direct' ']' 'sum' \\ \langle INCEPT_B \rangle &::= '[' \langle PRL_CONV_NS \rangle ']' 'concat' \\ &\quad | '[' \langle PRL_CONV_NS_RD \rangle ']' 'concat' \\ \langle INCEPT_RESN_B \rangle &::= '[' '[' \langle PRL_CONV_NS \rangle ']' 'concat' ',' 'direct' ']' 'sum' \\ \langle PRL_CONV_NS \rangle &::= \langle CONV_N_S1 \rangle ',' \langle CONV_N_S1 \rangle \\ &\quad | \langle PRL_CONV_NS \rangle ',' \langle CONV_N_S1 \rangle \\ \langle PRL_CONV_NS_RD \rangle &::= \langle CONV_N_S1 \rangle \langle RD \rangle ',' \\ &\quad \langle CONV_N_S1 \rangle \langle RD \rangle \\ &\quad | \langle CONV_N_S1 \rangle \langle RD \rangle ',' \langle RD \rangle \\ &\quad | \langle PRL_CONV_NS_RD \rangle ',' \langle CONV_N_S1 \rangle \langle RD \rangle \end{aligned}$$

$\langle RD \rangle ::= \langle CONV_S2 \rangle$
 $\quad | \quad \langle POOL_S2 \rangle$
 $\langle CONV \rangle ::= \langle CONV_Sk \rangle \mid \dots \forall k \in cpS$
 $\langle CONV_Sk \rangle ::= 'conv-i-j-k \ 1' \mid \dots \forall i \in cD, \forall j \in cF, \forall l \in norm$
 $\langle POOL \rangle ::= 'p-m-k' \mid \dots \forall p \in pT, \forall m \in pD, \forall k \in cpS$
 $\langle POOL_Sk \rangle ::= 'p-m-k' \mid \dots \forall p \in pT, \forall m \in pD, \forall k \in cpS$
 $\langle DROPOUT \rangle ::= 'dropout-r' \mid \dots \forall r \in dR$
 $\langle FC_B \rangle ::= 'fc-s' \mid \dots \forall s \in fS$

2.2. Diversity Control via Structural Schema Clusters

Due to the high number of parameters used by the expression language to define the architectures, measuring diversity by checking for unique expressions of population individuals can be misleading. By taking those expressions and ignoring their specific operator parameters, we can group them into clusters of unique abstract structural schemas. Those clusters represent groups of solutions where the connection structure between generic operations is the same. To give a simple example, both these two sub-expressions, “ap-3-1 [conv-3-128-1 dropout-0.4 , direct] sum conv-3-256-2 dropout-0.5” and “ap-3-1 [conv-3-56-1 dropout-0.2 , direct] sum conv-3-128-1 dropout-0.3”, belong to this generic structural schema: “ap [conv dropout , direct] sum conv dropout”.

The number of schema clusters gives us an idea of how many structurally different architectures are being explored by the current population.

2.3. Schema Diversity Factor Mutation

We propose a new grammar-based mutation operator based on Grammar Based Mutation (GBM) [73]. Since GBM only chooses one non-terminal node from the tree for mutation and uses a fixed mutation probability, we propose some modifications to adapt it to our NAS problem. We refer to our variant of GBM as Schema Diversity Factor (SDF) mutation and it differs from GBM in two main aspects: the derivation tree non-terminal nodes are checked for mutation following a width first search and the mutation probability (mP) is calculated for each particular individual derivation tree based around its non-terminal node number and the population diversity (d) following Equations (1)–(3).

$$mP(i) = \beta \cdot \frac{1}{count(i.nonTerminalNodes)} \quad (1)$$

where mP is the mutation probability for the individual i and β is a scaling factor that can be set manually or calculated in relation to the population diversity. We use the population structural diversity (d) measured via unique structural schema clusters with Equations (2) and (3).

$$\beta = (1 - d) \cdot (\beta_M - \beta_m) + \beta_m. \quad (2)$$

$$d = numberOfUniqueSchemasClusters / populationSize \quad (3)$$

Values of β belong to the interval $[\beta_m, \beta_M]$. During our experiments, the interval was set to $[0.5, 2]$.

This has the effects of increasing the mutation probability when the structural diversity is low in order to improve exploration and lowering it when said diversity is high to allow exploitation. In addition, since we check each node individually for mutation, more than one subtree at a time is allowed to mutate (more than one part of the same network architecture can be changed). Finally, we want a lower probability as the total node number grows to avoid an excessive number of mutations, which is why we adjust it for each individual depending on the number of non-terminal nodes.

2.4. Fitness

To obtain an individual fitness, its architecture expression is automatically translated into code and the resulting model is trained from scratch on the training partition of the dataset. For loss, we use

either the Root Mean Square Error (RMSE) or the Categorical Cross-Entropy (CCE) over the validation partition to obtain the fitness using the following equation:

$$fitness(indv) = \frac{1}{1 + \frac{1}{n} \cdot \sum_{i=lastTrainEpoch-(n-1)}^{lastTrainEpoch} loss(indv, validation, i)}$$

We use $n = 3$ to smooth the oscillations of the loss value and obtain an average estimation.

3. Experiments and Results

3.1. Experimental Setup

Two different servers were used to conduct the evolution tests (Table 1).

Table 1. Test servers: hardware specifications.

Server	GPUs	CPU	RAM
IGN1	4 × Tesla V100 16GB	2 × Intel(R) Xeon(R) Gold 6148@2.40GHz	128 GB
Cartobot1	2 × RTX 2080Ti 11GB	i7-8700 3.2GHz	64 GB

We restrict the search space of valid architectures for our experiments as follows:

- Maximum number of fully connected layers was set to 1 in order to reduce the memory and computational cost of training the network during the evolutive design stage. The maximum number of neurons on those layers was kept small for the same reason (8, 16, 32).
- Only convolutions of dimensions (1, 2, 3) were allowed, according to findings in [45] regarding factorized convolutions. The number of features is a set of typical used values in the range [2,256].
- We allowed the use of the two most frequently used types of pooling (max-pooling and average-pooling).
- Regarding the stride of convolutions and poolings, we used 1 to keep spatial dimensions and 2 to allow their reduction at certain parts of the architecture (those places are specified inside the meta-grammar restrictions).
- For the dropout rates, valid values for the drop-rate parameter were set inside the interval [0, 0.5] with increments of 0.1. Here, 0.0 is interpreted as an inactive dropout operator. The higher value 0.5 was set empirically to avoid TensorFlow-Keras warnings obtained for values above 0.5.
- We allowed the use of no normalization operation or Batch Normalization (BN) to let the design process choose where to place the BN operations.
- On all cases during design, the network was trained for its evaluation only for a maximum of 20 epochs (early stop if during 5 there was no improvement on the training metrics).

For the models labeled with the prefix “v2” (version 2), we allowed the following settings (due to higher computational capabilities of the server used for the experiments):

- Maximum number of FC layers was 3 instead of 1.
- Bigger dimensions for FC layers (up to 256).
- Separable 2D convolutions (“sepconv”) were used in addition to regular convolutions.
- BN momentum was empirically set to 0.5 (due to average size of training batches caused by GPU number and available memory).

On all of the experiments, the dataset was partitioned by experiment into train/validation/test subsets with a ratio of 70:20:10. This data partitioning was due to internal intermediate reports being filled with test metrics from that subset. Despite k-fold being used over the whole dataset, for final evaluation at the design step, we used only 90% of the dataset; we report this statistic for full

disclosure. Tournament selection used size $k = 4$ (*selectionPressure* took values 0.7 or 1 depending on the experiment).

3.2. Results

The dataset consists of 17,159 samples of 256×256 -pixel RGB images, labeled to indicate whether they contain roads (53.5%) or not (46.5%). The images are from the National Plan for Aerial Orthophotometry (PNOA) with original pixel resolutions of 0.25 or 0.50 m. They were taken from a Web Map Tiled Service (OGC:WMTS) of the IGN at zoom level 18 which corresponds to a scale of 1:2.132 so the equivalent pixel size on the ground is 0.597 m. The samples correspond to eight areas of the Spanish geography that can be considered representative: Albacete, Balearic Islands, Ciudad Real, Galicia, Huelva, Murcia, Navarre and Segovia. We refer to this dataset as IGN-PNOA hereafter.

Sample results of a designed network's output for three random sample groups (not used for training or validation) outside of the dataset (orthophotos were from random geographical areas of Spain with no associated cartography available) to check their predictions over unknown data are shown in Figure 4. Under each image of the group is the network predictive output for the class "contains road". In all these examples, the somewhat extreme output values given by the network are due to the use of the softmax activation on the output layer.

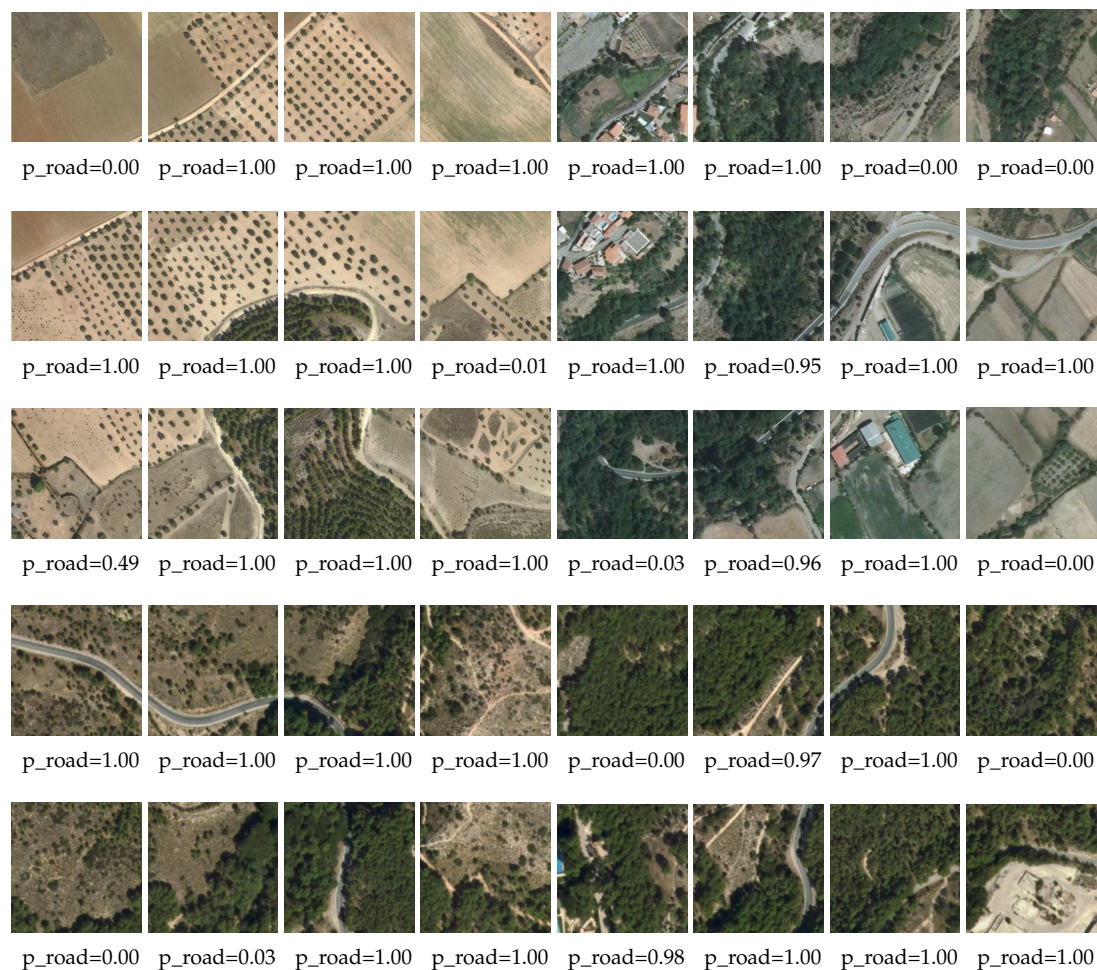


Figure 4. Sample predictions for three randomly selected zones outside the dataset that were not used for training. "p_road" shows the network prediction confidence for the class "contains road".

We selected the nine highest scoring networks (measured on the validation partition) designed by VGNet; all models were trained from scratch using the train and validation examples. To evaluate

their test accuracy, k-fold cross-validation (CV) was used with typical values of $k = 5, 10$ [81,82]. Regarding $k = 10$, the authors of [81,83] showed that 10-fold CV obtain similar results to leave-one-out cross-validation (LOOCV) and has lower computational cost. Test metrics are included in Figures 5 and 6 and detailed in Table 2. Their respective accuracy (95% confidence intervals, using the calculation method detailed by Caffo [84]) are in Table 3 to show the expected accuracy for each model. In all cases, the same parameters used for designing the networks ($epochs = 200$, $alpha = 1e^{-4}$, $optimizer = Adam$ and batch size calculated for each model based on GPU memory available) were used to train them on this final stage. Finally, for the two best scoring models in Table 4, additional metrics were calculated on the test partition: precision, recall, F1 score and Area Under the ROC-Curve (AUC). To give an example of one of the designed networks, a high-level architectural overview of model “v2m2” is shown in Figure 7 with sequences of operations grouped into blocks for clarity (as detailed in the figure legend). This particular model contains two Inception-ResNet-like sections at the areas with parallel operator branches. Regarding execution times, on average k-fold ($k = 5$) evaluation takes 52.8 h on IGN1 server ($4 \times$ Tesla V100 GPUs).

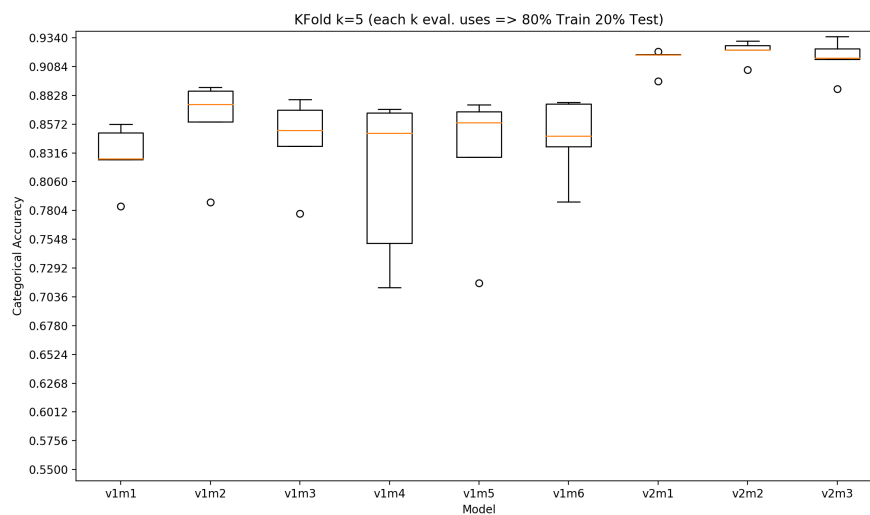


Figure 5. K-fold ($k = 5$) test accuracy for selected models on the IGN-PNOA dataset.

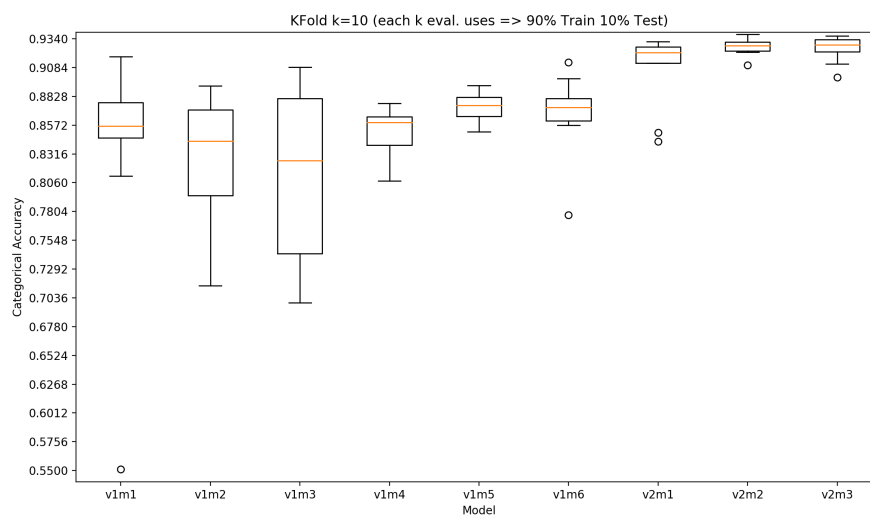


Figure 6. K-fold ($k = 10$) test accuracy for selected models on the IGN-PNOA dataset.

Table 2. K-fold test accuracy results for best designed networks (selected by validation accuracy) for the IGN-PNOA dataset.

Model	k	Accuracy Min	Accuracy Max	Accuracy Avg	Accuracy Stdev
v1m1	5	0.784	0.857	0.829	0.0284
	10	0.551	0.918	0.833	0.103
v1m2	5	0.788	0.890	0.860	0.042
	10	0.714	0.892	0.825	0.058
v1m3	5	0.777	0.879	0.843	0.040
	10	0.699	0.909	0.812	0.077
v1m4	5	0.712	0.871	0.810	0.073
	10	0.808	0.877	0.852	0.021
v1m5	5	0.716	0.874	0.829	0.0657
	10	0.851	0.892	0.874	0.012
v1m6	5	0.788	0.877	0.845	0.036
	10	0.777	0.913	0.867	0.0362
v2m1	5	0.895	0.922	0.915	0.011
	10	0.843	0.931	0.907	0.033
v2m2	5	0.906	0.931	0.922	0.010
	10	0.910	0.938	0.927	0.007
v2m3	5	0.889	0.935	0.916	0.017
	10	0.900	0.937	0.925	0.012

Table 3. Test accuracy (95% confidence intervals) for all models calculated from k-fold results for the IGN-PNOA dataset. For each model and k value, we show the calculated accuracy, expected interval mean, standard deviation and the interval itself.

Model	Test Accuracy 95% t Confidence Interval (k-fold k = 5)	Test Accuracy 95% Confidence Interval (k-fold k = 10)
v1m1	0.823 ± 0.035 = [0.793, 0.864]	0.833 ± 0.074 = [0.759, 0.907]
v1m2	0.860 ± 0.052 = [0.808, 0.912]	0.826 ± 0.042 = [0.784, 0.867]
v1m3	0.843 ± 0.050 = [0.793, 0.893]	0.812 ± 0.055 = [0.757, 0.867]
v1m4	0.810 ± 0.091 = [0.719, 0.901]	0.852 ± 0.015 = [0.836, 0.867]
v1m5	0.829 ± 0.082 = [0.747, 0.911]	0.874 ± 0.009 = [0.865, 0.882]
v1m6	0.845 ± 0.045 = [0.800, 0.890]	0.867 ± 0.026 = [0.842, 0.893]
v2m1	0.915 ± 0.013 = [0.901, 0.928]	0.907 ± 0.024 = [0.884, 0.931]
v2m2	0.922 ± 0.012 = [0.910, 0.934]	0.927 ± 0.005 = [0.922, 0.932]
v2m3	0.916 ± 0.021 = [0.894, 0.937]	0.925 ± 0.008 = [0.917, 0.934]

Table 4. Additional metrics for the two best models on the IGN-PNOA dataset on the test partition.

Model	Precision	Recall	F1 Score	A.U.C.
v2m2	0.936	0.904	0.920	0.950
v2m3	0.931	0.884	0.907	0.952

Test accuracies for some well-known architectures on the IGN-PNOA dataset obtained by members of the Cartobot project team [41] by applying transfer learning techniques (in this case using partitioning train/validation/test with ratio 50:25:25) are for Inception-v3 (87.88–91.19%) or Inception-ResNet (91.28–92.53%). In addition, in [43], the mean accuracies for the same dataset are reported, among others, for ResNet-50 (91.7%), VGG (93.9%) and for two different ensembles (94.6% and 95.6%, respectively). Here, ensemble refers to the combination of multiple networks into one meta-classifier, namely their results are combined to improve predictions.

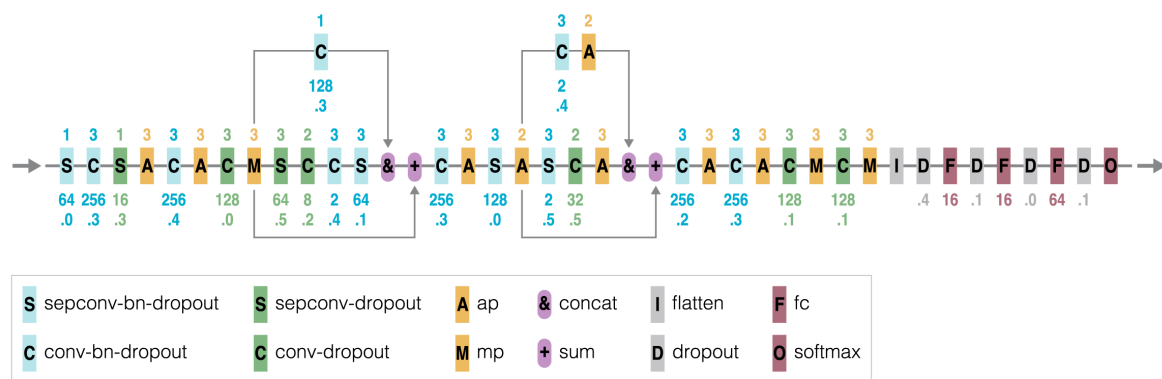


Figure 7. Example of a network automatically designed by VGNet. High-level architectural overview of model “v2m2”. Numbers indicate from top to bottom: (i) “S” and “C” blocks: size, number of filters and dropout rate; (ii) “A” and “M” blocks: size; (iii) “D” dropout rate; and (iv) “F” number of units.

Finally, to give a point of common reference for other authors to be able to compare results, we ran preliminary tests (with no additional fine-tuning of any kind) on a typical NAS benchmark, the CIFAR-10 dataset (we set the maximum number of FC layers to 3 on this experiments). Here, the best model (selected by validation loss) designed by VGNet obtained test categorical accuracies that oscillate in the range shown in Table 5. The gap between this results and the current state-of-the-art for CIFAR-10 can be explained by the fact that the grammar, as well as other internal hyperparameters, have been designed for road-detection on the IGN-PNOA dataset (including as mentioned some parameters that are set due to our current hardware constraints). This suggests that application to other problem domains, such as the case of CIFAR-10 or other remote sensing tasks, may need further grammar modifications and fine-tuning to achieve optimal results.

Table 5. Test metrics for best VGNet’s designed model for the CIFAR10 dataset.

Model	Accuracy Mean	Accuracy Stdev	Accuracy Mean, with Data Augmentation	Accuracy Stdev, with Data Augmentation
c10g49m2	0.911	0.003	0.918	0.002

4. Discussion and Future Work

Our system was able to successfully design full networks that, trained from scratch (without any manual fine-tuning or data augmentation or application of ensembles), yielded for the best scoring model a test accuracy of 90.6–93.1% with kfold ($k = 5$) and 91–93.8% (with $k = 10$). Therefore, the GGGP NAS approach shows results similar to well-known architectures. The main disadvantage we find is the time it takes to run the evolutive algorithm. We should note that, as mentioned above, some of the meta-grammar parameters were set in our tests according to hardware constraints (e.g., GPU memory) and therefore should not be interpreted as optimal settings in any case. Exploring ways of efficiently incorporating transfer-learning and data augmentation, fine tuning the system and adapting the grammar for pixel-level semantic labeling architectures are some of the next steps of the Cartobot project. On the other hand, meta-grammars allow new applications of the system as a hypothesis checker using different variations of meta-grammars and comparing the results obtained. For example: Is it better to apply dropout only at a certain stage of the architecture? What is the optimal value for the dropout rate?

We are currently working on methods to speed up the design process, from reducing the number of epochs at early design stages to using caching methods for the convolutional block weights or migrating to training on cloud servers. Research is also being conducted on ways to gain reusable knowledge among different evolution experiments or problem domains. Finally, to improve our

method's capability of balancing exploration/exploitation during the NAS, future research will further explore enhancements to the presented SDF mutation operator.

Author Contributions: Conceptualization, V.d.l.F.C., A.D.-Á. and F.S.G.; methodology, V.d.l.F.C., A.D.-Á. and F.S.G.; software, V.d.l.F.C.; validation, V.d.l.F.C., A.D.-Á. and F.S.G.; formal analysis, V.d.l.F.C.; investigation, V.d.l.F.C.; resources, V.d.l.F.C., A.D.-Á., M.-Á.M.-C. and F.S.G.; data curation, V.d.l.F.C.; writing—original draft preparation, V.d.l.F.C.; writing—review and editing, V.d.l.F.C., A.D.-Á., M.-Á.M.-C. and F.S.G.; visualization, V.d.l.F.C.; supervision, F.S.G.; project administration, M.-Á.M.-C.; and funding acquisition, F.S.G. and M.-Á.M.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Instituto Geográfico Nacional (IGN, Spain).

Acknowledgments: The authors acknowledge the staff at the Instituto Geográfico Nacional (IGN, Spain) and all team members of the Cartobot project; Carolina Temprado for her support and additional graphic design; and José Eugenio Naranjo for additional draft reviewing and suggestions.

Conflicts of Interest: The founders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Risojević, V.; Momić, S.; Babić, Z. Gabor descriptors for aerial image classification. In *International Conference on Adaptive and Natural Computing Algorithms*; Springer: London, UK, 2011; pp. 51–60.
2. Topouzelis, K.; Papakonstantinou, A.; Doukari, M. Coastline change detection using Unmanned Aerial Vehicles and image processing technique. *Fresen. Environ. Bull.* **2017**, *26*, 5564–5571.
3. Korzeniowska, K.; Bühler, Y.; Marty, M.; Korup, O. Regional snow-avalanche detection using object-based image analysis of near-infrared aerial imagery. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1823–1836. [\[CrossRef\]](#)
4. Yuan, C.; Ghamry, K.; Liu, Z.; Zhang, Y. Unmanned aerial vehicle based forest fire monitoring and detection using image processing technique. In Proceedings of the 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), Nanjing, China, 12–14 August 2016.
5. Zacharie, M.; Fuji, S.; Minori, S. Rapid Human Body Detection in Disaster Sites Using Image Processing from Unmanned Aerial Vehicle (UAV) Cameras. In Proceedings of the 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Bangkok, Thailand, 21–24 October 2018.
6. Selim, S.; Sonmez, N.; Coslu, M.; Onur, I. Semi-automatic tree detection from images of unmanned aerial vehicle using object-based image analysis method. *J. Indian Soc. Remote Sens.* **2019**, *47*, 193–200. [\[CrossRef\]](#)
7. Buters, T.; Belton, D.; Cross, A. Seed and seedling detection using unmanned aerial vehicles and automated image classification in the monitoring of ecological recovery. *Drones* **2019**, *3*, 53. [\[CrossRef\]](#)
8. Alidoost, F.; Arefi, H. A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **2018**, *86*, 235–248. [\[CrossRef\]](#)
9. Dutta, T.; Sharma, H.; Vellaiappan, A.; Balamuralidhar, P. Image analysis-based automatic detection of transmission towers using aerial imagery. In *Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Cham, Switzerland, 2015.
10. Tragulnuch, P.; Chanvimaluang, T.; Kasetkasem, T.; Ingprasert, S.; Isshiki, T. High Voltage Transmission Tower Detection and Tracking in Aerial Video Sequence using Object-Based Image Classification. In Proceedings of the 2018 International Conference on Embedded Systems and Intelligent Technology International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, 7–9 May 2018.
11. Lu, J.; Ma, C.; Li, L.; Xing, X.; Zhang, Y.; Wang, Z.; Xu, J. A Vehicle Detection Method for Aerial Image Based on YOLO. *J. Comput. Commun.* **2018**, *6*, 98–107. [\[CrossRef\]](#)
12. Cao, Y.; Wang, G.; Yan, D.; Zhao, Z. Two algorithms for the detection and tracking of moving vehicle targets in aerial infrared image sequences. *Remote Sens.* **2016**, *8*, 28. [\[CrossRef\]](#)
13. Malof, J.; Bradbury, K.; Collins, L.; Newell, R.G.; Serrano, A.; Wu, H.; Keene, S. Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier. In Proceedings of the 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 20–23 November 2016.
14. Yu, J.; Wang, Z.; Majumdar, A.; Rajagopal, R. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule* **2018**, *2*, 2605–2617. [\[CrossRef\]](#)

15. Shorter, N.; Kasparis, T. Automatic vegetation identification and building detection from a single nadir aerial image. *Remote Sens.* **2009**, *1*, 731–757. [[CrossRef](#)]
16. Fortier, A.; Ziou, D.; Armenakis, C.; Wang, S. *Survey of Work on Road Extraction in Aerial and Satellite Images*; Technical Report; Center for Topographic Information Geomatics: Ottawa, ON, Canada, 1999; Volume 241.
17. Wang, J.; Qin, Q.; Gao, Z.; Zhao, J.; Ye, X. A new approach to urban road extraction using high-resolution aerial image. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 114. [[CrossRef](#)]
18. Mattyus, G.; Wang, S.; Fidler, S.; Urtasun, R. Enhancing road maps by parsing aerial images around the world. In Proceedings of the IEEE International Conference on Computer Vision, Araucano Park, Chile, 7–13 December 2015; pp. 1689–1697.
19. Tsochantaridis, I.; Joachims, T.; Hofmann, T.; Altun, Y. Large margin methods for structured and interdependent output variables. *J. Mach. Learn. Res.* **2005**, *6*, 1453–1484.
20. Zhou, H.; Kong, H.; Wei, L.; Creighton, D.; Nahavandi, S. Efficient road detection and tracking for unmanned aerial vehicle. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 297–309. [[CrossRef](#)]
21. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [[CrossRef](#)]
22. Lin, Y.; Saripalli, S. Road detection from aerial imagery. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 3588–3593.
23. Unsalan, C.; Sirmacek, B. Road network detection using probabilistic and graph theoretical methods. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4441–4453. [[CrossRef](#)]
24. Hu, J.; Razdan, A.; Femiani, J.C.; Cui, M.; Wonka, P. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4144–4157. [[CrossRef](#)]
25. Hinz, S.; Baumgartner, A. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 83–98. [[CrossRef](#)]
26. Trinder, J.C.; Wang, Y. Automatic road extraction from aerial images. *Digit. Signal Process.* **1998**, *8*, 215–224. [[CrossRef](#)]
27. Airault, S.; Ruskone, R.; Jamet, O. Road detection from aerial images: A cooperation between local and global methods. In *Image and Signal Processing for Remote Sensing*; International Society for Optics and Photonics: Bellingham, WA, USA, 1994; Volume 2315, pp. 508–518.
28. Zlotnick, A.; Carnine, P. Finding road seeds in aerial images. *CVGIP Image Underst.* **1993**, *57*, 243–260. [[CrossRef](#)]
29. Bastani, F.; He, S.; Abbar, S.; Alizadeh, M.; Balakrishnan, H.; Chawla, S.; Madden, S.; DeWitt, D. Roadtracer: Automatic extraction of road networks from aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4720–4728.
30. Mátyus, G.; Luo, W.; Urtasun, R. Deeproadmapper: Extracting road topology from aerial images. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3438–3446.
31. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *Int. J. Remote Sens.* **2015**, *36*, 3144–3169. [[CrossRef](#)]
32. Mnih, V.; Hinton, G.E. Learning to detect roads in high-resolution aerial images. In *European Conference on Computer Vision*; Springer: London, UK, 2010; pp. 210–223.
33. Hu, X.y.; Zhang, Z.x.; Zhang, J. An approach of semiautomated road extraction from aerial image based on template matching and neural network. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 994–999.
34. Ichim, L.; Popescu, D. Road detection and segmentation from aerial images using a CNN based system. In Proceedings of the 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4–6 July 2018; pp. 1–5.
35. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [[CrossRef](#)]
36. Saito, S.; Aoki, Y. Building and road detection from large aerial imagery. In *Image Processing: Machine Vision Applications VIII*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; Volume 9405, p. 94050K.
37. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [[CrossRef](#)]

38. Rao, Y.; Liu, W.; Pu, J.; Deng, J.; Wang, Q. Roads detection of aerial image with FCN-CRF model. In Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2018.
39. Wei, Y.; Wang, Z.; Xu, M. Road structure refined CNN for road extraction in aerial image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 709–713. [\[CrossRef\]](#)
40. Zhong, Z.; Li, J.; Cui, W.; Jiang, H. Fully convolutional networks for building and road extraction: Preliminary results. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1591–1594.
41. Cira, C.I.; Alcarria, R.; Manso-Callejo, M.Á.; Serradilla, F. Evaluation of Transfer Learning Techniques with Convolutional Neural Networks (CNNs) to Detect the Existence of Roads in High-Resolution Aerial Imagery. In *Applied Informatics*; Florez, H., Leon, M., Diaz-Nafria, J.M., Belli, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 185–198.
42. Cira, C.I.; Alcarria, R.; Manso-Callejo, M.Á.; Serradilla, F. A Deep Convolutional Neural Network to Detect the Existence of Geospatial Elements in High-Resolution Aerial Imagery. *Proceedings* **2019**, *19*, 17. [\[CrossRef\]](#)
43. Cira, C.I.; Alcarria, R.; Manso-Callejo, M.Á.; Serradilla, F. A Framework Based on Nesting of Convolutional Neural Networks to Classify Secondary Roads in High Resolution Aerial Orthoimages. *Remote Sens.* **2020**, *12*, 765. [\[CrossRef\]](#)
44. Volpi, M.; Tuia, D. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [\[CrossRef\]](#)
45. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
46. Zoph, B.; Le, Q.V. Neural Architecture Search with Reinforcement Learning. *arXiv* **2016**, arXiv:1611.01578.
47. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. *arXiv* **2017**, arXiv:1707.07012.
48. Pham, H.; Guan, M.Y.; Zoph, B.; Le, Q.V.; Dean, J. Efficient Neural Architecture Search via Parameter Sharing. *arXiv* **2018**, arXiv:1802.03268.
49. Liu, H.; Simonyan, K.; Vinyals, O.; Fernando, C.; Kavukcuoglu, K. Hierarchical Representations for Efficient Architecture Search. *arXiv* **2017**, arXiv:1711.00436.
50. Real, E.; Aggarwal, A.; Huang, Y.; Le, Q.V. Regularized Evolution for Image Classifier Architecture Search. *arXiv* **2018**, arXiv:1802.01548.
51. Liang, J.; Meyerson, E.; Hodjat, B.; Fink, D.; Mutch, K.; Mikkilainen, R. Evolutionary Neural AutoML for Deep Learning. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2019), Prague, Czech Republic, 13–17 July 2019.
52. Assunção, F.; Lourenço, N.; Machado, P.; Ribeiro, B. DENSER: deep evolutionary network structured representation. *Genet. Program. Evolvable Mach.* **2019**, *20*, 5–35. [\[CrossRef\]](#)
53. Lima, R.H.R.; Pozo, A.T.R. Evolving Convolutional Neural Networks through Grammatical Evolution. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 179–180. doi:10.1145/3319619.3322058. [\[CrossRef\]](#)
54. Couchet, J.; Manrique, D.; Porras, L. Grammar-guided neural architecture evolution. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*; Springer: London, UK, 2007; pp. 437–446.
55. Tsoulos, I.; Gavrili, D.; Glavas, E. Neural network construction and training using grammatical evolution. *Neurocomputing* **2008**, *72*, 269–277. doi:10.1016/j.neucom.2008.01.017. [\[CrossRef\]](#)
56. Jacob, C.; Rehder, J. Evolution of neural net architectures by a hierarchical grammar-based genetic system. In *Artificial Neural Nets and Genetic Algorithms*; Springer: London, UK, 1993; pp. 72–79.
57. Mahmoudi, M.T.; Taghiyareh, F.; Forouzideh, N.; Lucas, C. Evolving artificial neural network structure using grammar encoding and colonial competitive algorithm. *Neural Comput. Appl.* **2013**, *22*, 1–16. [\[CrossRef\]](#)
58. Assunção, F.; Lourenço, N.; Machado, P.; Ribeiro, B. Automatic generation of neural networks with structured Grammatical Evolution. In Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastián, Spain, 5–8 June 2017; pp. 1557–1564.

59. Soltanian, K.; Tab, F.A.; Zar, F.A.; Tsoulos, I. Artificial neural networks generation using grammatical evolution. In Proceedings of the 2013 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, 14–16 May 2013; pp. 1–5.
60. Ahmadizar, F.; Soltanian, K.; AkhlaghianTab, F.; Tsoulos, I. Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm. *Eng. Appl. Artif. Intell.* **2015**, *39*, 1–13. [[CrossRef](#)]
61. Ahmad, Q.; Rafiq, A.; Raja, M.A.; Javed, N. Evolving MIMO Multi-Layered Artificial Neural Networks Using Grammatical Evolution. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 8–12 April 2019; pp. 1278–1285. [[CrossRef](#)]
62. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, Morgan Kaufmann Publishers, Inc.: Burlington, MA, USA, 2012; pp. 1097–1105.
63. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
64. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–10 June 2015; pp. 1–9.
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
66. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI* **2017** *4*, 12.
67. Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. Dual path networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4467–4475.
68. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–27 July 2017; pp. 5987–5995.
69. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (IJCV)* **2015**, *115*, 211–252. [[CrossRef](#)]
70. Whigham, P.A. Grammatically-based genetic programming. *Proc. Workshop Genet. Program. Theory-Real-World Appl.* **1995**, *16*, 33–41.
71. García-Arnau, M.; Manrique, D.; Rios, J.; Rodríguez-Patón, A. Initialization method for grammar-guided genetic programming. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*; Springer: London, UK, 2006; pp. 32–44.
72. Koza, J.R.; Koza, J.R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Cambridge, MA, USA, 1992; Volume 1.
73. Couchet, J.; Manrique, D.; Ríos, J.; Rodríguez-Patón, A. Crossover and mutation operators for grammar-guided genetic programming. *Soft Comput.* **2007**, *11*, 943–955. [[CrossRef](#)]
74. Goldberg, D.E.; Deb, K. *A Comparative Analysis of Selection Schemes Used in Genetic Algorithms*; FOGA: Washington, DC, USA, 1990.
75. Hingee, K.; Hutter, M. Equivalence of probabilistic tournament and polynomial ranking selection. In Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; pp. 564–571. [[CrossRef](#)]
76. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
77. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28*; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2015; pp. 91–99.
78. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
79. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.

80. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML* **2013**, *30*, 3.
81. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: London, UK, 2013; Volume 112.
82. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: London, UK, 2013; Volume 26.
83. Molinaro, A.M.; Simon, R.; Pfeiffer, R.M. Prediction error estimation: A comparison of resampling methods. *Bioinformatics* **2005**, *21*, 3301–3307. [[CrossRef](#)] [[PubMed](#)]
84. Caffo, B. *Statistical Inference for Data Science*; Leanpub: Victoria, BC, Canada, 2015.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).