

This is a postprint version of the following published document:

González-Díaz, Iván (2019). DermaKNet: Incorporating the knowledge of dermatologists to Convolutional Neural Networks for skin lesion diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 23(2), pp.: 547-559.

DOI: <https://doi.org/10.1109/JBHI.2018.2806962>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

See <https://www.ieee.org/publications/rights/index.html> for more information.

# DermaKNet: Incorporating the knowledge of dermatologists to Convolutional Neural Networks for skin lesion diagnosis

Iván González-Díaz, *Member, IEEE*,

**Abstract**—Traditional approaches to automatic diagnosis of skin lesions consisted of classifiers working on sets of hand-crafted features, some of which modeled lesion aspects of special importance for dermatologists. Recently, the broad adoption of Convolutional Neural Networks (CNNs) in most computer vision tasks has brought about a great leap forward in terms of performance. Nevertheless, with this performance leap, the CNN-based Computer Aided Diagnosis (CAD) systems have also brought a notable reduction of the useful insights provided by hand-crafted features. This paper presents DermaKNet, a CAD system based on CNNs that incorporates specific subsystems modeling properties of skin lesions that are of special interest to dermatologists, aiming to improve the interpretability of its diagnosis. Our results prove that the incorporation of these subsystems not only improves the performance, but also enhances the diagnosis by providing more interpretable outputs.

**Index Terms**—Skin lesion analysis, Melanoma, Convolutional Neural Networks, Dermoscopy, CAD

## I. INTRODUCTION

EARLY Melanoma Diagnosis is one of the traditional fields of application of Computer Aided Diagnosis (CAD) systems. In addition to the high incidence and aggressiveness of melanoma (it is the skin cancer that causes the most deaths in Europe [1]), there are other aspects that make it an specially suitable field for automatic diagnosis methods. For example, the early removal of the lesion completely cures the disease, effectively preventing metastasis [2]. Melanocytes are one of the very few cells that are naturally colored and visible to the eye, which make them possible to diagnose using clinical images. Also, the use of portable and affordable acquisition instruments, such as dermatoscopes, improves the accuracy in the diagnosis can be improved by 5-30% [3]. As a result, there is a growing interest in incorporating automatic systems in the daily practice of dermatologists, aiming not to replace their diagnosis, but to improve it by providing valuable information about the clinical case, and serving as filtering tools that automatically detect those cases with a high confidence of benignity, which can have a great impact in the final amount of moles that must be analyzed by the clinicians.

However, despite the research efforts devoted to the topic, these systems have yet to become part of everyday clinical practice. From our point of view, there are two factors currently hampering the adoption of CAD systems by

dermatologists. Firstly, the lack of large, open, annotated datasets, containing images of lesions gathered by different medical institutions and a great variety of dermatoscopes, has undermined the generalization capability of developed CAD systems, leading to poor results when applied to different datasets. Additionally, it has prevented standard and fair comparisons between proposed methods, thus hindering the scientific advances in the field. Secondly, most of CAD systems simply provide a tentative diagnosis to the clinicians, which does not actually help them much in practice. Hence, it would be more desirable for these systems to be able to provide some insight about the elements and properties of the lesion that support the diagnosis.

In regard to the first factor, the International Skin Imaging Collaboration: Melanoma Project (ISIC<sup>1</sup>) is an academia-industry partnership created to facilitate digital skin imaging technologies to help reduce melanoma mortality. In addition to developing standards to address the technologies, techniques, and terminology used in skin imaging, ISIC is continuously building an open source public access archive (ISIC Archive<sup>2</sup>) of skin images that allows researchers to assess and validate their CAD systems. The archive is large and includes images acquired using different devices from multiple medical institutions. Furthermore, since 2016, the association is also promoting the research in the field by organizing an International Challenge in which automatic methods for lesion segmentation, dermoscopic feature detection and skin disease diagnosis are evaluated using images of the archive [4] [5].

With respect to the second, in the last few years there have been changes in machine learning technology that have increased the difficulty of interpreting the results of the CAD systems. Whereas traditional approaches relied on low-level handcrafted features computed over the lesion [6] [7], some of them modeling aspects of special importance for dermatologists in their diagnosis [8] [9], modern approaches, such as [10] and [11], have adopted the use of Convolutional Neural Networks (CNNs), due to their impressive performance in many computer vision tasks such as classification [12] [13], detection [14] and segmentation [15] [16]. The drawback of CNN-based systems is the lack of clear understanding of the underlying factors and properties that support the final decision.

This paper presents DermaKNet (Dermatologist Knowledge

I. González-Díaz is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28045, Spain, e-mail: igonzalez@tsc.uc3m.es.

<sup>1</sup><http://isdis.net/isic-project/>

<sup>2</sup><https://isic-archive.com/>

Network), a CAD system for automatic diagnosis of skin lesions that aims to keep the best of both alternatives and, consequently, incorporates the intuitions of dermatologists into a CNN-based framework. By developing novel computational blocks in the net, we model properties of the lesions that are known to be discriminative for clinicians. The benefit of our approach is twofold: firstly, as we will demonstrate in the experimental section, the performance of the diagnosis is improved and, secondly, the interpretability of the system can be enhanced by analyzing the outputs of these expert-inspired blocks. In particular, our system includes several elements, not found in general-purpose classification CNNs, that become the main contributions of this work:

- A *Dermoscopic Structure Segmentation Network* which segments the lesion area into a set of high-level dermoscopic features corresponding to global and local structures that have turned out to be of special interest for dermatologists in their diagnosis. In the absence of strongly annotated data, we have trained this network from weakly-annotated clinical cases.
- A novel *Modulation Block* that incorporates these segmentations into the diagnosis process as probabilistic modulators of neuron activations.
- Two additional novel blocks, *Polar Pooling* and *Asymmetry*, that mimic the way in which dermatologists analyze skin lesions.
- *3-branch top-layers* in the diagnosis CNN, that provide the final diagnosis using both the traditional information channels as well as these novel pathways modeling expert intuitions.
- Some other elements with a great impact over the final system performance such as a specifically tailored data augmentation process, or an external classifier based on non-visual meta-data.

The remainder of this paper is organized as follows: Section II performs a review of the related literature. In Section III we provide a general description of our method for automatic diagnosis of skin lesions. Sections IV and V present our Dermoscopic Structure Segmentation and Diagnosis networks, respectively. Section VI explains the experiments and discusses the results that support our method and, finally, Section VII summarizes our conclusions and outlines future lines of research.

## II. RELATED WORK

Traditional approaches address the problem of automatic melanoma diagnosis using discriminative methods working over sets of hand-crafted visual features from dermoscopic images. These features vary from general-purpose descriptors, e.g. color and texture filter-banks [17] [18] [19], to problem-dependent knowledge-based features. The later deserve more interest from our point of view since they aim to model particular lesion aspects of special importance for dermatologists. Consequently, besides improving the system performance, they also enhance the interpretability of the automatic diagnosis [20]. In [8], the authors proposed a reduced set of interpretable features modeling some properties of the ABCD rule [21], such as symmetry and border sharpness. Along the same lines,

some other methods start by detecting a set of dermoscopic structures that are later used to generate the diagnosis. Examples of these dermoscopic features include reticular patterns, dots and globules, streaks, etc. The complete set of structures that are commonly considered was defined in the *pattern analysis method* for melanoma diagnosis [3], which has been widely adopted by specialists due to its accurate results.

To tackle the problem of detection of dermoscopic features, classical segmentation techniques, such as Gaussian Mixture Models [22], Markov Random Fields [23] and Topic Models [9], or even discriminative approaches working over textons [24], have been adopted in the literature. Once the areas corresponding to some of these structures have been identified, diagnosis can be inferred: in [25] the ABCD rule is combined with structure recognition in an attempt to detect suspicious lesions, in [26] the 7-point checklist method is applied to the outputs of these structure detectors, and in [9] probabilistic segmentation maps are used to build a set of specific classifiers, each one focusing on a particular structure, which are then fused to provide the final diagnosis.

During the last few years, with the advent and broad adoption of CNNs in many recognition problems in computer vision, several works have been proposed that apply this paradigm to melanoma classification. In [10] CNNs are combined with sparse coding and SVMs to provide a diagnosis. In [27] a Fully Convolutional Neural Network (FCNN) is first used to segment the input image into lesion area and surrounding skin; then a square and tight cropping is performed, and finally a diagnosis is provided using a CNN that is fine-tuned from the well-known resnet model [13]. In [28], the authors have trained a CNN using a very large dataset with 129,450 clinical images and 2,032 different diseases, and tested its performance against 21 board-certified dermatologists on biopsy-proven clinical images with two critical binary classification use cases: malignant carcinomas versus benign seborrheic keratoses and malignant melanomas versus benign nevi. Their results show that the automatic system achieves similar performance than all tested experts across both tasks, demonstrating a level of competence comparable to dermatologists. However, despite their impressive performance when enough training data is available, CNN-based methods still lack a clear understanding of the underlying factors and properties that support their final decision, limiting their usability and preventing their broad adoption by dermatologists.

In this paper, we propose to incorporate knowledge-based interpretable properties of skin lesions into the framework of CNNs. Although Majtner et al. [29] have previously tried to fuse hand-crafted features with CNNs, their approach simply fused the outputs of two independent classifiers (one based on hand-crafted features and the other using a CNN) to generate the final diagnosis. To the best of our knowledge, this is the first attempt to achieve a seamless integration between the knowledge of dermatologists and CNNs. For that purpose, we have developed several novel processing blocks, with the dual goal of improving the system performance and gaining interpretability in the diagnosis.

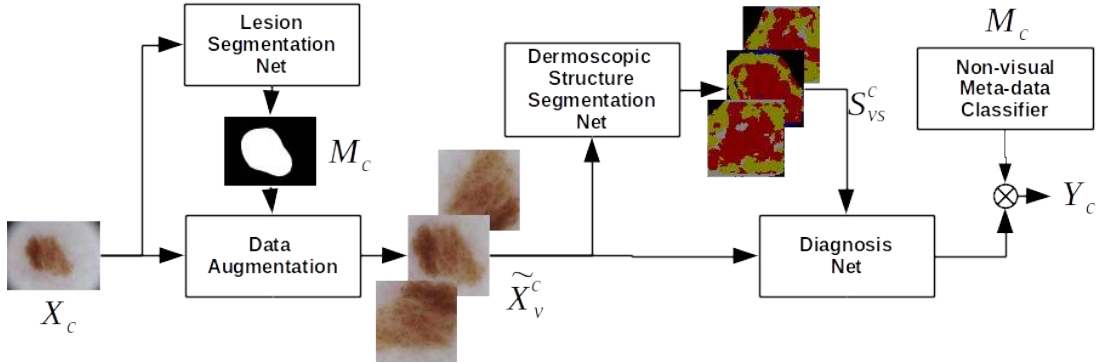


Fig. 1. Main processing pipeline of DermaKNet. Each clinical case is defined by an image  $X_c$ . The Lesion Segmentation Net firstly segments the image into areas corresponding to lesion and surrounding skin, giving rise to the binary masks  $M_c$ . Then, the Data Augmentation Module extends the initial visual support of the lesion and generates additional views  $\tilde{X}_v^c$  of the lesion by applying rotations and crops. Next, the Dermoscopic Structure Segmentation Network segments each lesion view into a set of high level dermoscopic structures  $s$ . Finally, the whole set of the lesion images  $\tilde{X}_v^c$  and their corresponding segmentation maps  $S_{vs}^c$  are passed to the Diagnosis Network, which generates the diagnosis.

### III. AN AUTOMATIC METHOD FOR SKIN LESION DIAGNOSIS

In this section we will provide a general description of our CAD system and also explain those processing blocks that, although have an important impact in the system performance, do not constitute the main contributions of our paper. Finally, these main contributions will be later described in their own sections.

#### A. General description of the system

The main pipeline of DermaKNet is depicted in Fig. 1. It comprises the following steps:

1) For each clinical case  $c$ , a dermoscopic image  $X_c$  is first passed to the *Lesion Segmentation Network (LSN)*, which generates a binary mask  $M_c$  outlining the area of the image corresponding to the lesion. A description of this module is given in Section III-B.

2) Next, the pair  $\{X_c, M_c\}$  goes through the *Data Augmentation Module*. This module extends the initial visual support of the lesion and generates additional views  $v$  of the lesion by applying rotations and crops. Hence, the output of this module is an extended set of images  $\tilde{X}_v^c$  representing the clinical case. Section III-C provides a detailed description of the data augmentation process.

3) The following step in the process is performed by the *Dermoscopic Structure Segmentation Network (DSSN)*. It aims to segment each view of the lesion  $\tilde{X}_v^c$  into a set of eight dermoscopic features corresponding to global and local structures that have turned out to be relevant for dermatologists in their daily practice. Examples of these structures are dots/globules, regression areas, streaks, etc. The output of this subsystem is a set of 8 segmentation maps  $S_{vs}^c$ ,  $s = 1..8$ , each associated with one of the considered structures. This module, as well as the format of the segmentation maps, will be introduced in Section IV.

4) The augmented set  $\{\tilde{X}_v^c, S_{vs}^c\}$  is passed to the *Diagnosis Network (DN)*, which provides a tentative diagnosis for the clinical case. The description of this network can be found in Section V.

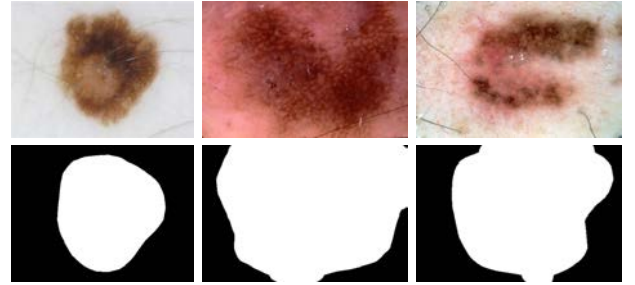


Fig. 2. Some examples of clinical cases (top) and the binary lesion segmentations computed by our Lesion Segmentation Module (bottom).

5) If additional non-visual meta-data about the lesion (e.g. patient age, sex, etc.) are available, the previous diagnosis is further factorized using the score of a classifier working over these non-visual information to produce the final diagnosis  $Y_c$ . This classifier is described in Section III-D.

#### B. Lesion Segmentation Network (LSN)

The Lesion Segmentation Network (LSN) has been developed by training a Fully Convolutional Network (FCN) [15]. FCNs have achieved state-of-the-art results on the task of semantic image segmentation (general content), as demonstrated in the PASCAL VOC Segmentation task [30]. In order to train a network for our particular task of lesion-skin segmentation, we have used the training set for the lesion segmentation task in the 2017 ISBI challenge [5].

In Figure 2 we show various examples of lesion segmentations computed by this module. Let us note that the goal is not to generate very accurate segmentation maps, but to produce binary masks  $M_c$  that broadly identify the area of the image that corresponds to the lesion.

#### C. Data Augmentation Module and Normalized Polar Coordinates

It is well known that data augmentation often boosts the performance of deep neural networks, mainly when the amount of available training data is limited. Among all the potential

image variations and artifacts, invariance to orientation is probably the main requirement in our particular scenario, as dermatologists do not follow a specific protocol during the acquisition of an image with the dermatoscope. More complex geometric transformations such as affine or projective transformations are less interesting since the dermatoscope is normally placed just over and orthogonally to the lesion surface.

Based on these observations, the particular process of data augmentation for a given clinical case  $c$  is illustrated in Fig. 3 and described next:

1) First, starting from the pair  $\{X_c, M_c\}$ , we generate a set of rotated versions (see Fig. 3b).

2) Since rotating an image without losing any visual information requires adding new areas that did not exist in the original view, we find and crop the largest inner rectangle satisfying that all pixels belong to the original image (Fig. 3c). We have observed that removing these black areas (using the inner rectangle) at the expense of losing some regions of the lesion produces better results than keeping the whole lesion and the black regions. The rationale behind is that, although some of lesion details may be partially lost in some of the views, we are considering all of them by analyzing the whole augmented set of lesion views. However, this fact makes necessary to perform data augmentation also in test.

3) Finally, since our subsequent CNNs (Structure Segmentation and Diagnosis) require square input images of  $256 \times 256$  pixels, we perform various squared crops which are in turn re-sized to these dimensions (see Fig. 3d-3f).

In particular, in our approach we have considered 8 rotations and 3 crops for each clinical case  $c$ , leading to an augmented set of 24 images, each one represented by a tensor  $\tilde{X}_v^c \in \mathbb{R}^{256 \times 256 \times 3}$ , with  $v = 1 \dots 24$ .

In addition, and during the data augmentation process, we compute the pixel *Normalized Polar Coordinates* for each generated view  $\tilde{X}_v^c$ . The goal of these coordinates is to provide invariance against shifts, rotations, changes in size and even irregular shapes of the lesions in subsequent processing steps. To do so, we transform the original Cartesian pixel coordinates  $(x_i, y_i)$  into the normalized polar coordinates  $(r_i, \theta_i)$ , where  $r_i \in [0, 1]$  and  $\theta_i \in [0, 2\pi)$  stand for the normalized ratio and angle, respectively. The process that computes this transformation is as follows: first, the binary mask of the lesion is approximated by an ellipse with the same second-order moments. Then, we learn an affine matrix  $A$  that transforms the ellipse into a normalized (unit ratio) circle centered at location  $(0,0)$ . Finally, we map each pixel in the original lesion with its projection in the normalized circle, and obtain the normalized polar coordinates as the ratio and angle computed from the projected pixel coordinates. Figure 4 shows an example of a rotated and cropped view of a lesion and its corresponding normalized polar coordinates.

#### D. Considering non-visual lesion meta-data in the diagnosis

As we will describe in detail in Section VI, the proposed model has been used to participate in the 2017 ISBI Challenge on Skin Lesion Analysis Towards Melanoma Detection. In

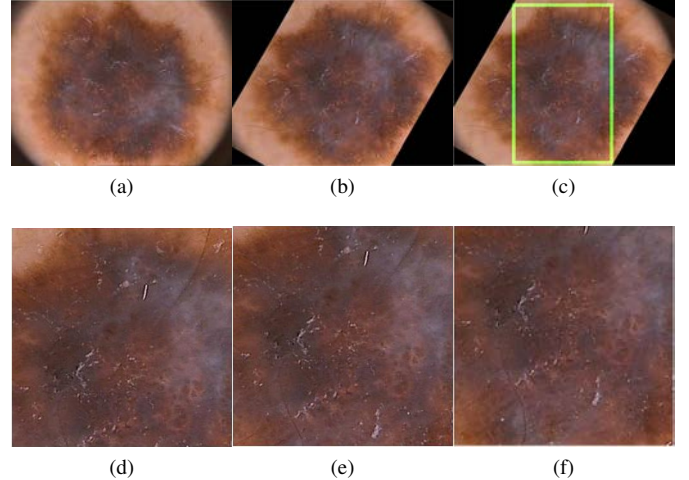


Fig. 3. Illustration of the process of data augmentation: a) Original image, b) rotated Image, c) largest inner rectangle containing pixels of the lesion, d-f) 3 square croppings containing partial views of the lesion.

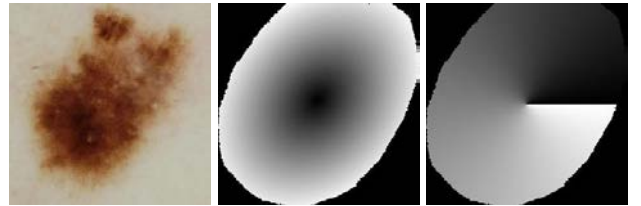


Fig. 4. Example of a rotated and cropped view of a lesion and its Normalized Polar Coordinates. (Left) View of the lesion. (Middle) Values of the normalized ratio. (Right) Angle.

this challenge, additional valuable meta-data was also provided with the images that can help automatic systems to improve their performance, namely: a) approximate age of the patient, rounded to 5 year intervals (or ‘unknown’ if not available), and b) sex, containing the gender of the patient (or ‘unknown’ if not available).

Hence, we have complemented the outputs of the CNN-based system with the score provided by a Support Vector Machine (SVM) [31] working on external non-visual meta-data. Since both age and sex variables are discrete, we have transformed them into numerical inputs following this process: for each considered category  $c$  and the meta-data  $M$ , we have modeled the corresponding likelihoods  $p(M|c)$  as random discrete variables. Then, given a new clinical case, the input to the SVM was computed by evaluating the likelihood of the current sample.

Furthermore, we have also included an additional input feature computed as the relative area of the lesion with respect to the total size of the image.

As shown in Figure 1, a probabilistic output of this SVM is then factorized with the output of the Diagnosis Network to provide the final diagnosis  $Y_c$  of the system.

#### IV. DERMOSCPIC STRUCTURE SEGMENTATION NETWORK (DSSN)

The goal of the Dermoscopic Structure Segmentation Network is the following: given an input view of the lesion  $\tilde{X}_v^c$



it aims to provide a segmentation considering a pre-defined set of dermoscopic features that correspond with global and local structures of special interest for dermatologists in their diagnosis.

#### A. Considered Dermoscopic Structures

In this work we have considered a set of eight structures:

1.- *Dots, globules and Cobblestone pattern* [32, pp. 15-17]: although different, they have been fused into one for the purpose of the system development due to their visual similarities. These patterns consist of a certain number of round or oval elements, variously sized, with shades that can be brown and gray-black. In the case of cobblestone structures, they are usually larger, more densely grouped and somewhat angulate. In general, they are often located in lesion areas that are growing. While an even spatial distribution with regular size and shape is associated with benignity, various sizes and shapes, or irregular or localized distribution usually occur in melanoma. Depending on their relative extent in the lesion area, these features can be either identified as local structures or global patterns.

2.- *Reticular pattern and pigmented networks* [32, pp. 10-13]: they cover most parts of certain lesions. They look as grids of thin brown lines over a light brown background and are quite common in melanocytic lesions. If globally distributed, this structure is related to benign lesions. However, variations in size and form are indicative of malignancy. Depending on their relative extent in the lesion area, these features can be either identified as local structures or global patterns.

3.- *Homogeneous areas* [32, pp. 14-15]: these areas are diffuse, with brown, grey-black, grey-blue or reddish-black shade, where there is no other local feature that can be recognized. A globally distributed pattern of bluish hue is the hallmark of the blue nevus. With other shades, it may be present in several types of lesions, such as Clark-nevi, dermal nevi or nodular and metastatic melanomas. Depending on their relative extent in the lesion area, these features can be either identified as local structures or global patterns.

4.- *Regression* [32, pp. 20-21]: these structures are generally well-defined white and/or blue areas that appear when the immune system has attacked the lesion. White areas resemble a superficial scar, and blue areas may appear as diffuse blue-gray areas or peppering, which is an aggregation of blue-gray dots. Regression areas are always considered local structures.

5.- *Blue-white veil* [32, pp. 22-23]: a region of grey-blue to whitish-blue blurred pigmentation, correlated with pigmented network disorder (globules or streaks) and highly indicative of melanoma. This structure is always considered local in our annotations.

6.- *Streaks* [32, pp. 17-18]: are black or light to dark brown longish structures of variable thickness, not clearly combined with pigmented networks, and easily observed when located at the periphery of the lesion. In general, they tend to converge to the center of the lesion. An even, radial distribution of the streaks around the border of the lesion is characteristic of Reed nevus. However, an asymmetric or localized distribution of streaks suggests malignancy. This structure is always local, and spatially localized on the lesion borders.

7.- *Vascular structures* [32, p. 23]: they are homogeneous areas with vessels. Depending on their shape, they may be a clear sign of malignancy. While abundant and prominent comma vessels often exist in dermal nevi, some other vascular patterns, such as arborizing, hairpin or linear irregular ones, are more frequent in melanomas. These structures are always considered local in our annotations.

8.- *Unspecific pattern*: we group in this category those parts of the lesion that cannot be assigned to any of the previous structures. No direct diagnosis implication can be inferred from it. Nevertheless, it is more often related to melanoma, or at least it suggests that the lesion must be carefully explored. Depending on its relative extent in the lesion area, these feature can be either identified as local or global in our annotations.

#### B. A weak learning approach for segmentation

The main challenge to develop the DSSN is the annotation of the training dataset. A traditional supervised approach would require to provide a ground truth pixel-wise segmentation for each training image. This kind of strong annotation is often hard to obtain as it demands a huge effort from the dermatologists to manually outline the segmentations of the structures. Alternatively, providing weak image-level labels indicating only which dermoscopic structures are present in a lesion is much easier for dermatologists and becomes more affordable. Henceforth, following this alternative approach, we asked dermatologists of a collaborating medical institution, the *Hospital Doce de Octubre* in Madrid, to annotate the *ISIC 2016 training dataset* [4] with the presence or absence of the 8 aforementioned dermoscopic structures. In particular, we asked them to provide one label  $L(s)$  per structure  $s$  and clinical case:  $L(s) = 0$  if the structure is not present,  $L(s) = 1$  if it takes up just a local area of the lesion (*local structure*),  $L(s) = 2$  if it is present and dominant enough to be considered a *global pattern* in the lesion.

Given this weakly-annotated dataset, we have developed a segmentation network based on the method described in [33], where the authors introduced a Constrained Convolutional Neural Network for weakly supervised segmentation. For the sake of completeness we will include here some equations of the original model that accommodate the extensions and modifications for our particular scenario. For an in-depth discussion and derivation of these equations, the interested reader is referred to the original paper [33].

To keep the notation simple, we omit the image index in the following paragraphs. Let us consider the dermoscopic structure segmentation as a pixel-wise labeling problem in which each pixel  $i$  in the lesion area is labeled as belonging to a particular structure  $s_i$ ,  $s = 1 \dots P = 8$  or to a background class ( $s = 0$ ). Passing the input image through the segmentation CNN will produce a spatially reduced score map  $f_i(s_i; \theta)$  ( $64 \times 64$  in our case) at its top layer, where  $\theta$  represents the set of parameters of the CNN. Applying a parametric *softmax* over the network scores, we can model the label of each pixel location  $i$  as a probabilistic random variable with value  $q_i(s_i|\theta)$ :

$$q_i(s_i|\theta) = \frac{1}{Z_i} \exp(\gamma f_i(s_i|\theta)) \quad (1)$$

Here  $s_i$  is the random variable that represents the label (dermoscopic structure) at the location  $i$  and  $Z_i = \sum_{s=0\dots P=8} \exp(\gamma f_i(s|\theta))$  is the partition function at the location  $i$ . The utility and appropriateness of the parameter  $\gamma$ , which was not included in the original model, will be discussed later on.

At this point, we have introduced another modification to the original model. Our problem is highly unbalanced and, whereas structures as dots/globules or reticular patterns are very common, others like blue-white veil or regression patterns are less frequent. Moreover, the frequency of a structure does not correspond with its impact on the diagnosis, and less frequent patterns are in general more indicative of malignant lesions. We have observed that learning the model directly from the data leads to solutions that focus more on the correct segmentation of the most frequent patterns, while fail in those less-frequent but more meaningful patterns. To avoid such a situation, we have introduced weights that control the influence of the different structures in the learning process and produce more balanced segmentations. Hence, considering marginal independence, the probability distribution on an image can be factorized as:

$$Q(S|\theta) = \prod_i^N q_i(s_i|\theta)^{w_i} \quad (2)$$

where  $N$  is the total number of pixels in the spatially-reduced image generated by the CNN. The weights  $w_i$  might simulate the repetition of a sample in the training data, thus giving it more influence over the learned model. Although these weights can control the influence of each pixel  $i$  in the image, in our case, the value is the same for all pixels in the image  $w_i = w$  and represents a measure of the lesion abnormality, which depends on the weak ground-truth labels indicating the presence or absence of each dermoscopic structure. In fact, this weight  $w$  is inversely proportional to the likelihood of the present structures; if  $p^+(s)$  is the probability that a lesion contains the structure  $s$ , we compute  $w$  as:

$$w = \frac{1}{P} \sum_{s=1}^P 1[L(s) > 0]p^-(s) + 1[L(s) = 0]p^+(s) \quad (3)$$

where  $p^-(s) = 1 - p^+(s)$ , and  $1[\cdot]$  is an indicator function which is evaluated only when the inner condition is satisfied.

Given the probability distribution of an image stated in (2), the constrained CNN optimization for weakly-supervised segmentation proposed in [33] is:

$$\begin{aligned} & \text{find} && \theta \\ & \text{subject to} && A\vec{Q} \geq \vec{b} \end{aligned} \quad (4)$$

where  $\vec{Q}$  is the vectorized form of the network output  $Q(S|\theta)$ , and  $A \in \mathbb{R}^{K \times PN}$  and  $\vec{b} \in \mathbb{R}^K$  define  $K$  linear constraints over the output distribution  $Q$ . Since this problem is not convex with respect to the network parameters  $\theta$ , the authors defined a variational latent probability distribution  $P(S)$  over the semantic labels, which is independent of the CNN parameters  $\theta$ , applied the constraints to this new distribution rather than to the original network output  $Q(S|\theta)$ , and enforced  $P(S)$  and  $Q(S|\theta)$  to model the same probability distribution

by minimizing the Kullback-Leibler divergence between them. The resulting formulation becomes a Lagrangian optimization problem and gives rise to the following update equation:

$$p_i(s) = \frac{1}{Z_i} \exp(\gamma f_i(s_i; \theta) + A_{i,s_i}^T \lambda) \quad (5)$$

where  $\lambda \geq 0$  are the dual variables introduced in the optimization, and  $Z_i = \sum_s \exp(\gamma f_i(s; \theta) + A_{i,s}^T \lambda)$  is the local partition in location  $i$ . Additionally the final loss and its gradient needed by the optimization become:

$$\mathcal{L}(\theta) = - \sum_i \sum_{s_i} w_i p_i(s_i) \log q_i(s_i|\theta) \quad (6)$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial f_i(s_i)} = \gamma w_i [q_i(s_i|\theta) - p_i(s_i)] \quad (7)$$

The presence or absence of a dermoscopic structure, as well as additional cues about its size (global, local) or its location (borders, center) within the lesion, lead to particular constraints in the model. These constraints are applied over the accumulated probability  $P(s) = \sum_i p_i(s|\theta)$ , computed over all pixel locations in the segmentation map. In contrast to the original formulation in (4), we can now apply the constraints to the latent distribution  $A\vec{P} \geq \vec{b}$ , considering the following cases:

- *Absent structure*: If a dermoscopic structure  $s$  is not present in an image, we impose one constraint that acts as an upper bound over the accumulated probability:

$$\sum_{i=1}^N p_i(s) \leq 0 \quad (8)$$

- *Global structure*: If a dermoscopic structure  $s$  is considered as a global pattern, we impose one constraint acting as a lower bound over the accumulated probability  $P(s)$ , enforcing that a minimum area of the lesion corresponds to that structure:

$$l_s \leq \sum_{i=1}^N p_i(s) \quad (9)$$

Here  $l_s$  has been set to  $l_s = 0.5N$ , thus requiring that at least 50% of pixels in the lesion belong to that structure.

- *Local structure*: If a structure  $s$  is local, we impose two constraints acting as lower and upper bounds over the accumulated probability  $P(s)$ , respectively:

$$l_s \leq \sum_{i=1}^N p_i(s) \leq u_s \quad (10)$$

where  $l_s = 0.10N$  and  $u_s = 0.5N$ , are the lower and upper bounds.

- *Spatially localized structures*: since some of the structures tend to appear in particular locations of the lesion, we can enforce our model to learn this dependency. This is the case, for example, of the streak pattern, which only appears in the borders of the lesion. Hence, considering that the location of

a dermoscopic feature is restricted to certain region  $R$ , and defining  $\bar{R}$  as its complement, we impose two constraints:

$$\sum_{i \in \bar{R}} p_i(s) \leq 0 \quad (11)$$

$$l_s \leq \sum_{i \in R} p_i(s) \quad (12)$$

where  $l_s = 0.5N_R$ , being  $N_R$  the number of pixels in region  $R$ . We define the region  $R$  using the Normalized Polar Coordinates introduced in Section III-C, which allows us, for example, to define ring-shaped areas modeling the outer part of a lesion.

Once we have defined the constraints, we can discuss the role of the parameter  $\gamma$  in the *softmax* function (see eq. (1)). We have observed that using a simple non-parametric *softmax* function leads to situations in which constraints over local structures were often obeyed by simply assigning some residual probability to every pixel in the segmentation map. This residual probability is not enough to assign any pixel to the local structure (they show higher probabilities for other structures), but allows for fulfilling the constraints over the accumulated probability. From our point of view, this is an undesired behavior since what one would like to have instead is a small region of pixels with high probabilities of belonging to the corresponding local structure. In other words, we prefer pixels that are clearly associated with a particular class, as long as they produce an actual image segmentation, rather than pixels with some residual probability for each category. To address this issue, we use values of  $\gamma \geq 1$  so that we can control how the *softmax* approximates the *max* function while it remains differentiable. In our case, we have used a value of  $\gamma = 2$ .

We have implemented the DSSN taking the well-known *resnet-50* [13] as initialization, removing the top layers, and using the *ISIC 2016 training dataset* [4] and the described constrained optimization with weak annotations. This module produces, for each view  $v$  of a clinical case  $c$ , a tensor  $S_v^c \in \mathbb{R}^{64 \times 64 \times 8}$  that contains the 8 probability maps of the considered structures.

To sum up, we have extended the original approach in [33] with three contributions: we have introduced a parametric *softmax* that helps to model the problem constraints preventing certain malfunctions; we have incorporated instance weights to the problem statement so that we can deal with the unbalanced nature of labels; and, finally, we have extended the set of constraints by adding one new family that allows us to take advantage of the prior knowledge about the spatial location of structures in the lesion.

In Figure 5 we show some examples of the segmentation maps generated by the DSSN. Let us note that, just to provide a simplified visualization in this figure, we have transformed the tensor  $S_v^c$  containing eight probabilistic maps, into a hard segmentation in which each pixel has been assigned to the most likely category. However, no spatial post-processing techniques (such as Markov Fields or other smoothing algorithms) have been applied.

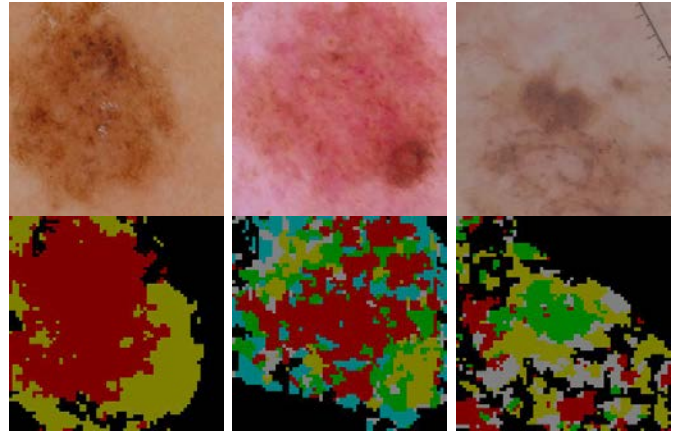


Fig. 5. Three illustrative results of the Dermoscopic Structures Segmentation Network. Top row: original images. Bottom row: segmentations. Colors represent dermoscopic structures: brown is dots/globules, mustard is reticular pattern/pigmented networks, green is homogeneous, grey is hypopigmented/regression areas, and blue is vascular structures. It is worth noting that, for the sake of easy visualization, each pixel has been assigned to the most probable category.

## V. DIAGNOSIS NETWORK (DN)

The *Diagnosis Network* (DN) gathers information from the previous modules and generates a diagnosis for each clinical case.

As in the previous module, we have also taken the *resnet-50* [13] as a basis, which uses residual layers to avoid the degradation problem when more and more layers are stacked to the network. When applied to our 256x256-pixel images, the last convolutional block (*res5c*) of this network produces a tensor  $T_c \in \mathbb{R}^{8 \times 8 \times 2048}$  containing the scores of high-level latent concept detectors (e.g. in Imagenet, the dataset for which it was originally designed, those were 2048 latent visual concepts).

In the original network, an average pooling layer transforms this tensor into a single-value per channel and image  $T_s \in \mathbb{R}^{1 \times 1 \times 2048}$ , which is followed by a fully connected layer and a *softmax* that generates the vector containing the probabilities of the considered visual concepts. Hence, the goal of the average pooling is to fuse detections at various locations of the input image and to generate a unified score for each latent high-level concept.

In our approach, we have modified the structure of the top layers of the network, giving rise to the pipeline illustrated in Figure 6. In the following sections, we will first introduce the structure of the top layers in the DN and, then, we will provide a detailed description of those blocks that have been specifically designed to work with dermoscopic images of skin lesions.

### A. Overview of the DN

As shown in the Figure 6, we have introduced several blocks to generate a final diagnosis  $Y_v^c$  for each considered view of a clinical case. Compared to the original *resnet-50*, we first apply a Modulation Block over the outputs of the convolutional *res5c* layer. This block, described in Section V-B, aims to modulate the previous outputs using the probabilistic segmentation maps



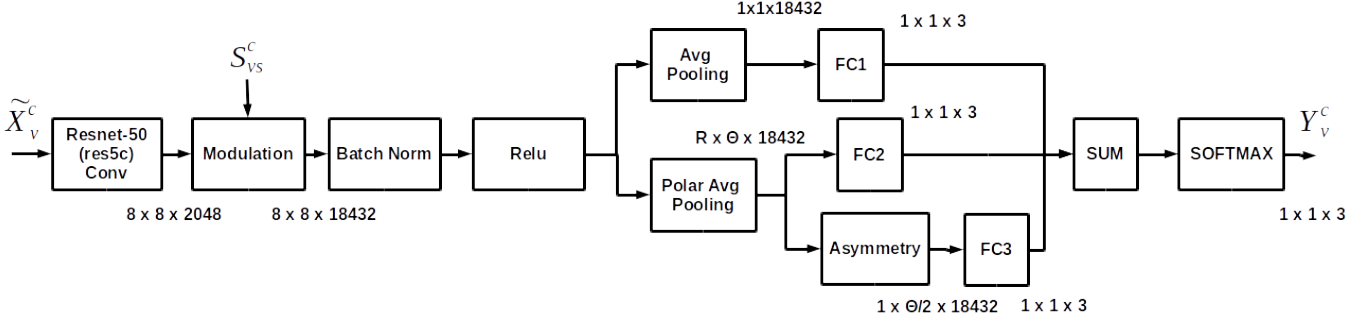


Fig. 6. Processing pipeline of the Diagnosis Network. The outputs of layer *res5c* of the original resnet-50 are modulated by the segmentation maps coming from DSSN, providing an extended set of channels. These channels, after batch normalization and ReLU activation function, are passed through a 3-branch processing pipeline that analyzes the presence of visual patterns, their spatial location, and the asymmetry of the lesion, respectively, to generate the diagnosis.

provided by DSSN. As explained below, this block multiplies the total number of channels or latent visual patterns by 9, which goes from 2048 to 18432. Next, the Modulation Block is followed by a Batch Normalization and non-linear ReLU activation (Rectified Linear Unit) layers. Finally, rather than just applying the Average Pooling + Fully Connected approach as in the original resnet-50, we have subdivided the pipeline into three parallel processing branches:

- 1) Branch 1: the original pipeline with an average pooling (8x8 in our case), followed by a fully connected layer (FC1).
- 2) Branch 2: it performs an average normalized polar pooling (see section V-C for further details) ( $R \times \Theta$ ;  $R = 3$ ,  $\Theta = 8$ ) followed by a fully connected layer (FC2). This branch provides a spatially discriminant analysis of the lesion.
- 3) Branch 3: it follows the previous polar pooling, estimates the asymmetry of the lesion (see Section V-D for a complete description), and applies a fully connected layer (FC3) over the asymmetry measures.

The outputs of these three branches are then linearly combined using a Sum Block, and the class-probabilities are computed using a *softmax*. Finally, in order to generate a unified final output for each clinical case  $Y_c$ , we consider independence between views leading to a factorization:

$$Y_c = \prod_{v=1}^V Y_v^c \quad (13)$$

### B. Modulation Block

The goal of the Modulation Block is to incorporate the segmentations provided by DSSN to the diagnosis process. To do so, this block fuses the structure segmentation maps described in Section IV-B with the outputs of the previous layer in the CNN.

In particular, if the output of the previous layer is a tensor  $\mathbf{x} \in \mathbb{R}^{M \times N \times O}$ , where  $M \times N$  are the dimensions of the output and  $O$  is the number of output channels, and  $\mathbf{s} \in \mathbb{R}^{M \times N \times P}$  is a segmentation map that has been previously re-sized to match the feature map, the output of this module is an extended and modulated feature map  $\mathbf{y} \in \mathbb{R}^{M \times N \times OP}$ . To compute this output, we modulate the  $o$ -th channel  $\mathbf{x}_o$  with the  $s$ -th segmentation map  $\mathbf{s}_s$ , producing a modulated channel  $\mathbf{y}_k$ :

$$\mathbf{y}_k = \mathbf{x}_i \odot \mathbf{s}_s, \quad i = 1 \dots O, s = 1 \dots P, k = 1 \dots OP \quad (14)$$

Since the segmentations computed by DSSN are fixed, this module has no parameters to be optimized during the training phase. Hence, the backpropagation process only requires the derivative with respect the data:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}_o} = \sum_{k \in K_o} \frac{\partial \mathbf{z}}{\partial \mathbf{y}_k} \odot \mathbf{s}_k \quad (15)$$

where  $K_o$  corresponds to all the modulated channels  $k$  generated from the channel  $o$ , and  $\mathbf{s}_k$  is the corresponding modulating map for that  $k$ .

The application of this module to our diagnosis network has been adapted as follows: it has been added to the network just after the *res5c* layer of the original resnet-50 [13]. Hence, we modulate  $O=2048$  channels using the probabilities of the  $P=8$  segmentation maps of local and global structures described in Section IV-A. In addition, we also concatenate the original input channels to the modulated ones, resulting into an extended set of  $O(P+1)$  channels (18432 in our case).

### C. Polar pooling

This block aims to perform pooling operations (average or max pooling), but instead of doing them over rectangular spatial regions, these operations are done over sectors defined in polar coordinates. Hence, for a given number of rings  $R$  (with  $r \in [0, 1]$ ) and angular sectors  $\Theta$  (angles  $\theta \in [0, 2\pi)$ ), this block transforms an input  $\mathbf{x} \in \mathbb{R}^{M \times N \times O}$  into an output  $\mathbf{y} \in \mathbb{R}^{R \times \Theta \times O}$ , where  $O$  is the number of channels.

Furthermore, in order to deal with lesions of irregular shape, we use the normalized polar coordinates described in Section III-C. Since, depending on the particular shape of a lesion and the size of the tensor being pooled, some combinations  $(r, \theta)$  may not contain pixels within the lesion, we can also define overlaps between adjacent sectors to improve the smoothness of the outputs. Moreover, we use a non-uniform radius quantization in order to generate fixed-area rings that contain the same number of pixels in the hypothetical case of an ideally circular lesion. To that end, the  $k$ -th ring is defined as:

$$\sqrt{\frac{k-1}{R}} \leq r < \sqrt{\frac{k}{R}} \quad (16)$$

for  $k = 1 \dots R$ . Given the proposed normalized polar coordinate system, the equations needed to perform the forward and

backward steps in the inference process do not differ from those ones of a regular max or average pooling block in Cartesian coordinates. Furthermore, it is worth noting that, once this block is applied and data is converted into polar coordinates, no more convolutional layers can be applied as the spatial relationships between contiguous values in the output matrix have been redefined (e.g. considering that columns in the data matrix refer to angles, the first and last columns are adjacent in the angular space). For that reason, in our approach, this module is followed by some blocks that are, either fully connected, or specifically designed to work with polar coordinates (e.g. the Asymmetry block).

#### D. Asymmetry Block

Melanomas tend to grow differently along each direction, becoming more asymmetric than benign lesions. This is why symmetry is present in a variety of diagnosis algorithms, such as the ABCD rule of dermoscopy [21]. The symmetry rule requires finding the axis of maximum symmetry according to some criteria (e.g. shape, color), and its perpendicular. In doing so, the lesion is labeled by dermatologists either as symmetric in one or two axes, or as asymmetric.

Our asymmetry block computes metrics that evaluate the asymmetry of a lesion with respect to various axes. In particular, given a polar division of the lesion into  $R \times \Theta$  sectors, we compute the asymmetry for axes aligned with the  $\Theta/2$  angles in the range  $[0, \pi)$ . To do so, our approach folds the lesion over each angle  $\theta$  and computes the accumulated square difference between corresponding sectors. Hence, for a given input  $\mathbf{x} \in \mathbb{R}^{R \times \Theta \times O}$ , this module generates an output  $\mathbf{y} \in \mathbb{R}^{1 \times \Theta \times O}$  as follows:

$$y_{\theta_k, o} = \frac{1}{R\Theta} \sum_{i=1}^R \sum_{j=1}^{\Theta/2} (x_{r_i, \theta_k + j} - x_{r_i, \theta_k - j})^2 \quad (17)$$

where, in case the angle index  $\theta_j$  becomes  $j <= 0$  it is substituted by  $\Theta - j$ .

During back-propagation, the gradients needed by the stochastic gradient descent algorithm are:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}_{r_i, \theta_j, o}} = \frac{2}{R\Theta} \sum_{k=1}^{\Theta/2} \frac{\partial \mathbf{z}}{\partial \mathbf{y}_{\theta_k, o}} \varphi(r_i, \theta_j, \theta_k) \quad (18)$$

where:

$$\varphi(r_i, \theta_j, \theta_k) = \begin{cases} x_{r_i, \theta_j} - x_{r_i, \theta_k - j}, & \theta_j \in [\theta_k, \theta_k + \pi) \\ x_{r_i, \theta_k - j} - x_{r_i, \theta_j}, & \text{otherwise} \end{cases} \quad (19)$$

#### E. Details about learning and evaluation processes

In this section we provide some useful details about the learning process of DN. As mentioned in Section V, we have taken the original resnet-50 as initialization and fine-tuned the network using our own training data. When nothing else is specified, all the new layers in the network are initialized using weights computed using Xavier's method [34].

Furthermore, due to the high degree of expressiveness of branches 2 and 3 with respect to the first branch, we have observed that training the whole system at a time was prone to

overfitting. Hence, instead, we have first trained a model using only the first branch with a learning rate of  $Lr = 10^{-4}$  and a weight Decay of  $Wd = 10^{-4}$ . Once a coarse convergence is reached, we have added the other two branches, frozen all layers up to (and including) the Modulation Block, initialized weights for branch 2 and 3 to zero, and learned the weights of the upper layers using the following learning rates:

- For branch 1 the original learning rate  $Lr_1 = 10^{-4}$  and weight Decay  $Wd = 10^{-4}$ .
- For branches 2 and 3 the original learning rate and weight decays are divided or multiplied by the total number of input spatial neurons in the fully connected block, respectively. This stronger regularization and slower learning rate prevents these branches from getting more relevance than the original one due to their expressiveness, and therefore minimizes the likelihood of overfitting.

The code that implements DermaKNet is available online<sup>3</sup>.

## VI. EXPERIMENTAL SECTION

### A. Datasets and Experimental Setup

DermaKNet has been assessed using the official dataset of the 2017 ISBI Challenge on Skin Lesion Analysis Towards Melanoma Detection<sup>4</sup> [5]. This challenge consists of three different parts: 1) Lesion Segmentation, 2) Detection and Localization of Visual Dermoscopic Features/Patterns, and 3) Disease Classification. We focus on part 3, being our goal the automatic diagnosis of dermoscopic images into three different categories: 1) *Nevus*: benign skin tumor, derived from melanocytes (melanocytic), 2) *Melanoma*: malignant skin tumor, derived from melanocytes (melanocytic), and 3) *Seborrheic Keratosis*: benign skin tumor, derived from keratinocytes (non-melanocytic). Malignancy diagnosis data were obtained from expert consensus and pathology report information.

The official dataset contains 2750 dermoscopic images gathered from the daily clinical practice of a wide range of medical centers, thus varying in resolution and capturing devices and conditions. The dataset has been split into training, validation and test sets with 2000, 150, and 600 images, respectively. The proportions of the classes in the training dataset are the following: 374 melanomas, 254 seborrheic keratosis, and 1372 benign nevi. Similar proportions were found in the test dataset, but not in the validation dataset. Besides the images, additional non-visual meta-data about the clinical cases such as the approximate age and sex of the patient were provided when available.

In addition to the official dataset, we have also considered two other external resources for training, namely:

- A dataset obtained from the EDRA Interactive Atlas of Dermoscopy [35], with images gathered with the objective of providing a panoramic view of skin lesions diagnosis using dermoscopy. Images were saved in JPEG format with almost uniform sizes (768x512), with ground truth labels determined by histopathological diagnosis, and no meta-data provided with the image files. It contains a total set of 724 melanocytic

<sup>3</sup><https://github.com/igondia/matconvnet-dermoscopy>

<sup>4</sup><https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a>

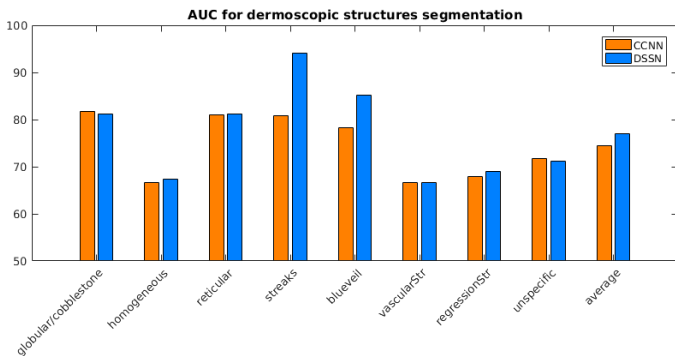


Fig. 7. AUC results for segmentation of dermoscopic structures. Our approach (DSSN) is compared to the original CCNN [33]

lesions, 222 of them are melanomas and the remaining 502 are benign nevi.

- A dataset built from the ISIC archive [36], the global repository that was used to generate the official challenge datasets. By considering images belonging to any of the three considered categories, we have found 2104 new images that were not included in the official dataset of the challenge: 1606 nevi, 466 melanomas and 32 seborrheic keratosis.

In order to assess the performance of our approach, we have used the same evaluation metrics proposed by the organizers of the challenge. In particular, among all the considered metrics, we will focus on:

- Area Under Curve (AUC): the AUC is computed independently for two different binary classification problems: 1) melanoma vs rest, and 2) seborrheic keratosis vs rest. In addition, the average AUC of the two problems is also provided. AUC was considered the main evaluation metric in the challenge and served to rank the official submissions and select the winning approaches.

- Specificity evaluated at a sensitivity of 95% (SP95): this complementary metric evaluates how automatic methods can filter out benign lesions, delivering to dermatologists only those that become good candidates to be malignant. Hence, by fixing a very high value on the sensitivity, we ensure that our system minimizes the number of non-detections, and assess its ability to reduce the clinical effort of dermatologists, which would have a high impact in their daily clinical practice.

The experiments in this section are organized as follows: first, we assess the performance of DSSN, then we evaluate the influence of each individual proposed block in the diagnosis network. Finally, the optimal configuration is evaluated in comparison to the official submissions of the challenge.

### B. Evaluation of the Dermoscopic Structure Segmentation Network

Prior to the analysis of the segmentation network, which is the final goal of this subsystem, we would like to provide an assessment of the DSSN module. To that purpose, we have subdivided our 2016 ISBI challenge training dataset with weak annotations into train and validation sets. Figure 7 shows a comparison between the original CCNN [33], which has served as a baseline for our method, and our approach

TABLE I

A COMPARISON BETWEEN SEVERAL VERSIONS OF DERMAKNET AND THE BASELINE. AUC IS GIVEN FOR MELANOMAS VS REST (MEL AUC), SEBORRHEIC KERATOSIS VS REST (SK AUC), AND AVERAGE (AVG AUC).

No	Method	Mel AUC	SK AUC	avg AUC
1	Resnet-50 [13]	83.1	91.8	87.4
2	Modulation + Branch 1	83.4	92.5	88.0
3	Modulation + Branches 1-3	83.6	92.7	88.2
4	(3) + Test Data Aug.	84.8	94.6	89.7
5	(4) + meta-data	<b>85.9</b>	<b>95.8</b>	<b>90.8</b>

including spatially localized structures, parametric *softmax* and instance weights that model lesion abnormality in training. Let us note that for the baseline CCNN we have considered three weak annotations (absent, local and global) using the same parameters as in our solution. Results are given in terms of AUC computed at image level, by accumulating the probability of the considered dermoscopic features over the pixels in the lesion. We can see in the figures that our proposal improves the performance of the baseline for some of the structures, specially on those that are spatially localized (e.g. streaks) or less frequent in the dataset (e.g. streaks and blueveil), in both cases as a consequence of our extensions in the model (spatially localized structures and training weights). In average, we are getting an improvement of 2.5 in AUC with respect to the baseline.

### C. Assessment of the proposed blocks in the automatic diagnosis system

In this set of experiments, we have assessed how the different blocks designed for our system help to improve the performance of the diagnosis. Although during the preparation of the challenge we have used the train and validation sets to make decisions about the system configuration, here we have preferred to show the results over the test set (see Table I). The rationale behind is that, as we have already mentioned, the validation dataset follows a particular data distribution, with different proportions of melanoma, keratosis and benign nevu than those found in the training and test sets. Hence, results in validation dataset, although show similar behaviors, are different in absolute terms and less meaningful for the analysis. In addition, it should be mentioned that these models have been trained using only the official 2017 ISBI training set.

Table I shows the results of this experiment. Let us note that, unless specified, the data augmentation described in section III-C is only performed over the training data. The first evaluated method is the original resnet-50 [13], which has been fine-tuned using our training data, and becomes the baseline algorithm in the comparison. The second approach substitutes the top layers in the original network by our Modulation Block (sec. V-B) followed by the regular Branch 1, and shows that incorporating the dermoscopic segmentations provides an absolute improvement of 0.60 in terms of average AUC. When we further incorporate the multi-branch processing with spatial and asymmetry analysis, we gain an additional increment of 0.2 in average AUC, which, although not very notable, is still valuable. By applying data augmentation also to the



test samples and fusing the results using the factorization introduced at the end of Section V-A, we get a quite significant additional improvement of 1.5 in average AUC. From our point of view, this improvement comes from the following fact: the inner rectangles during data augmentation remove the non-lesion black areas at the expense of also losing some areas of the lesion. This leads to training images that show partial views of the lesion, which requires to perform the same process for the test images in order to establish fair comparisons between data samples. Finally, the full system incorporating all the previous blocks as well as the score of a SVM classifier over the non-visual meta-data (see section III-D) achieves the best results, with an additional gain of 1.1 in average AUC. Hence, all these results demonstrate how each proposed extension enhances the quality of the diagnosis.

#### D. Comparison with the state-of-the-art

In this section we assess the performance of DermaKNet (version 5 in Table I) trained over an extended dataset containing both the official 2017 ISBI Challenge train dataset and the two external resources. In Table II we show a comparison between our method and the top 5 performing methods among the 23 official submissions to the challenge.

It is worth noting that we have included two versions of DermaKNet. Our official submission to the challenge, denoted as *DermaKNet (Official)* was trained using the official training dataset and the EDRA external resource and shows some minor differences with respect to the model described here (the interested reader is referred to [38] for the corresponding description). Our current proposal, denoted as *DermaKNet (Current)*, in contrast, was trained using also images from the ISIC archive.

Considering the official submission and using the AUC, DermaKNet ranked first in the category of Seborrheic Keratosis vs rest, and fourth in Melanoma vs rest, achieving a global second position in the challenge. However, if we further analyze the SP95 results, that account for the specificity of the diagnosis at a 95% sensitivity, our model clearly outperforms the rest of the approaches in both categories. As we have already mentioned, this probably represents the most realistic scenario of application, in which CAD systems are used to filter out benign cases, thus reducing the number of lesions that require care from dermatologists.

However, our current implementation achieves even better results than our previous one. Again, using AUC as the main performance metric, it outperforms all the official submissions in the Seborrheic Keratosis vs rest problem, and now also in the average of the two categories. Furthermore, our results for melanoma detection are now very close to those of the winning approach in the category. In addition, considering the SP95 metric, our method clearly becomes the state-of-the-art in the considered three problems (any of the two tasks and average). Hence, these results completely validate our approach and demonstrate the utility of incorporating intuitions from dermatologists into the CNN structure.

#### E. An example of an interpretable diagnosis

The advantage of incorporating the intuitions of dermatologists into the processing pipeline of a CNN is not restricted to the enhancement of the system performance, but also provides additional valuable information regarding the clinical case that might help medical staff in their diagnosis. Figure 8 shows two examples of a system output built using DermaKNet: we can provide information about the dermoscopic features, not only about their location in the lesion, but also about their contribution to the final diagnosis. In particular, the bottom-center diagram shows the accumulated contribution to the diagnosis score from each particular dermoscopic structure  $s$  (see eq. (14)). For that purpose, given the diagnosed class, we compute the non-probabilistic class-score of each dermoscopic feature  $s$  as the output of the Sum Block in Fig. 6 in which only those channels  $k$  ( $k = 1 \dots 18432$ ) that correspond with the structure  $s$  are considered in the computations. After applying a ReLU that removes negative (inhibiting) scores, we compute the final relative contribution by applying a normalization that ensures a total score of 1 over all the dermoscopic structures. Furthermore, the bottom-right diagram shows a normalized (for visualization enhancement) measure of the per-angle accumulated symmetry over the  $k$  channels modulated by each dermoscopic structure  $s$ . This symmetry is computed as follows: we consider the accumulated output of the asymmetry block for each angle  $y_{\theta_k} = \sum_o y_{\theta_k, o}$  (see eq. (17)), and then perform two consecutive normalizations: the first one adapts the asymmetry values to the lesion content dividing  $y_{\theta_k}$  by the accumulated energy of the input  $e = \sum_{r,j,o} x_{r_i, \theta_j, o}^2$  (see eq. (17)); and the second is a max-min normalization that ensures a final asymmetry in the range [0,1] and improves visualization. Finally, symmetry values are computed as 1-asymmetry.

In addition to gaining more insight on the automatic diagnosis, this information might also become the basis for other end-user applications, such as e-learning tools to help in the training of new specialists.

## VII. CONCLUSIONS AND FURTHER WORK

In this paper we have introduced DermaKNet, a CAD system for the diagnosis of skin lesions that is composed of several CNNs, each one devoted to a specific task: lesion-skin segmentation, detection of dermoscopic features, and global lesion diagnosis. Our goal through the whole system is to incorporate the expert knowledge provided by dermatologists into the decision process, overcoming the traditional limitation of deep learning regarding the lack of interpretability of the results. In order to achieve a seamless integration between CNNs and this expert information, we have developed several novel processing blocks.

We have assessed our system in the challenging dataset used in the *2017 ISBI Challenge on Skin Lesion Analysis Towards Melanoma Detection*, in the task of automatic diagnosis of melanoma and seborrheic keratosis. Our results prove that modeling expert-based information enhances the system performance and achieves very competitive results. In particular, the last version of our model ranks first in

TABLE II  
A COMPARISON BETWEEN DERMAKNET AND THE TOP FIVE PERFORMING OFFICIAL SUBMISSIONS TO 2017 ISBI CHALLENGE ON SKIN LESION ANALYSIS. AUC AND SP95 ARE GIVEN FOR MELANOMAS (MEL) VS REST, SEBORRHEIC KERATOSIS (SK) VS REST, AND AVERAGE (AVG).

Method	Mel AUC	SK AUC	avg AUC	Mel SP95	SK SP95	Avg SP95
Matsunaga et al. [37]	86.8	95.3	<b>91.1</b>	36.6	78.4	57.5
DermaKNet (Official) [38]	85.6	<b>96.5</b>	91.0	<b>40.4</b>	<b>82.4</b>	<b>61.4</b>
Menegola et al. [39]	<b>87.4</b>	94.3	90.8	39.5	69.0	54.3
Bi17 et al. [40]	87.0	92.1	89.6	39.8	47.6	43.7
Yang et al. [41]	83.0	94.2	88.6	36.6	74.5	55.6
DermaKNet (Current)	87.3	<b>96.2</b>	<b>91.7</b>	<b>46.0</b>	<b>84.3</b>	<b>65.2</b>

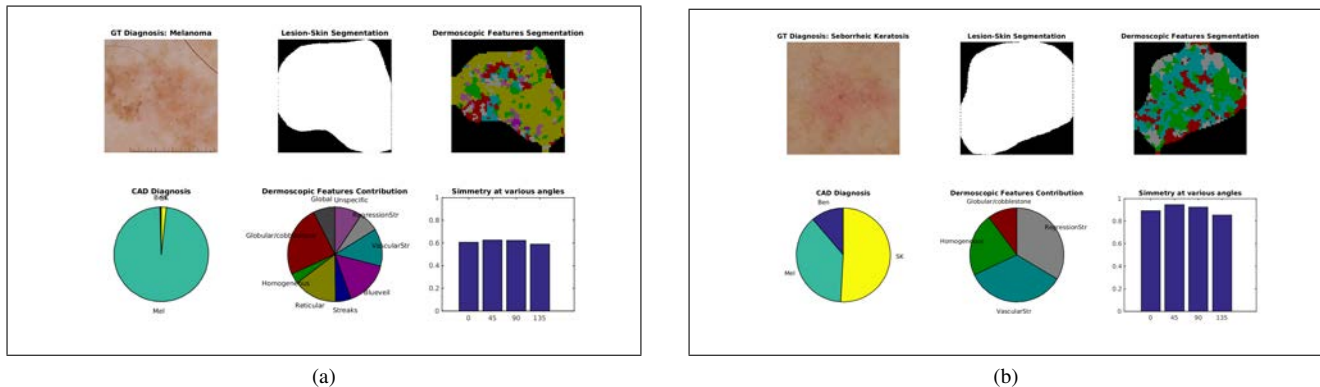


Fig. 8. Two examples of an interpretable system output generated with our approach: (a) Melanoma and (b) Seborrheic Keratosis. Each case contains 6 figures which represent (from top to bottom and left to right): original image and diagnosis, binary mask with lesion/skin, segmentation into dermoscopic features, automatic diagnosis, contribution of each dermoscopic feature to the final diagnosis, symmetry measures by angle.

the Seborrheic Keratosis category and average AUCs, and is very competitive in melanoma. Furthermore, our results in Specificity at a 95% Sensitivity are clearly better than those of the rest of the approaches, which makes our system very suitable as an automatic filtering module reducing the workload of dermatologists.

In addition to this gain in performance, we have also shown that we can produce a more interpretable diagnosis on top of our system. Looking at the outputs of those intermediate blocks modeling intuitions from dermatologists, we can get more insight about which dermoscopic features are influencing the diagnosis, the lesion symmetry, and even the spatial locations that support certain diagnosis.

The main lines of further research comprise the design of new blocks implementing other aspects in the lesions that are of interest for dermatologists, the development of segmentation methods that account for other useful dermoscopic features, and the exploration of novel ways of incorporating the dermoscopic structures segmentation into the diagnosis process. With respect to the latter, we will consider multi-task losses [42], which allow for sharing processing layers in both tasks and fusing segmentation and diagnosis networks into end-to-end trainable architectures.

#### ACKNOWLEDGMENTS

We kindly thank dermatologists of *Hospital 12 de Octubre of Madrid* because of their inestimable help annotating the data contents with the weak labels of structural patterns. This work was supported in part by the National Grant TEC2014-53390-P and National Grant TEC2014-61729-EXP of the Spanish

Ministry of Economy and Competitiveness. In addition, we gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN X GPU used for this research.

#### REFERENCES

- [1] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. Coebergh, H. Comber, D. Forman, and F. Bray, "Cancer incidence and mortality patterns in europe: Estimates for 40 countries in 2012," *European Journal of Cancer*, vol. 49, no. 6, pp. 1374 – 1403, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959804913000075>
- [2] M. A. Weinstock, "Cutaneous melanoma: public health approach to early detection," *Dermatologic Therapy*, vol. 19, no. 1, pp. 26–31, 2006. [Online]. Available: <http://dx.doi.org/10.1111/j.1529-8019.2005.00053.x>
- [3] H. Pehamberger, A. Steiner, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 17, no. 4, pp. 571 – 583, 1987. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0190962287702394>
- [4] D. Gutman, N. Codella, M. E. Celebi, B. Helba, M. Marchetti, N. K. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)," 05 2016.
- [5] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. K. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC)," *CoRR*, vol. abs/1710.05006, 2017. [Online]. Available: <http://arxiv.org/abs/1710.05006>
- [6] A. Madooei, M. S. Drew, M. Sadeghi, and M. S. Atkins, *Intrinsic Melanin and Hemoglobin Colour Components for Skin Lesion Malignancy Detection*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 315–322. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-33415-3\\_39](http://dx.doi.org/10.1007/978-3-642-33415-3_39)



- [7] P. Rubegni, G. Cevenini, M. Burrioni, R. Bono, P. Sbano, M. Biagioli, M. Risulo, N. Nami, R. Perotti, C. Miracco, and M. Fimiani, "Objective follow-up of atypical melanocytic skin lesions: a retrospective study," *Archives of Dermatological Research*, vol. 302, no. 7, pp. 551–560, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s00403-010-1051-6>
- [8] M. Zortea, T. R. Schopf, K. Thon, M. Geilhufe, K. Hindberg, H. Kirchesch, K. Mllersen, J. Schulz, S. O. Skrvseth, and F. Godtlielsen, "Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists," *Artificial Intelligence in Medicine*, vol. 60, no. 1, pp. 13 – 26, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0933366713001589>
- [9] J. López-Labraca, M. Á. Fernández-Torres, I. González-Díaz, F. Díaz-de María, and Á. Pizarro, "Enriched dermoscopic-structure-based cad system for melanoma diagnosis," *Multimedia Tools and Applications*, Jun 2017. [Online]. Available: <https://doi.org/10.1007/s11042-017-4879-3>
- [10] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, *Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images*. Cham: Springer International Publishing, 2015, pp. 118–126. [Online]. Available: [https://doi.org/10.1007/978-3-319-24888-2\\_15](https://doi.org/10.1007/978-3-319-24888-2_15)
- [11] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, April 2017.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 770–778. [Online]. Available: <https://doi.org/10.1109/CVPR.2016.90>
- [14] R. Girshick, "Fast r-cnn," in *International Conference on Computer Vision (ICCV)*, 2015.
- [15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1605.06211, 2016. [Online]. Available: <http://arxiv.org/abs/1605.06211>
- [16] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Cham: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [17] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1–2, pp. 61–81, 2005.
- [18] M. Abedini, Q. Chen, N. Codella, R. Garnavi, and X. Sun, ser. *Digital Imaging and Computer Vision*. CRC Press, Sep 2015, ch. Accurate and Scalable System for Automatic Detection of Malignant Melanoma, pp. 293–343, 0. [Online]. Available: <http://dx.doi.org/10.1201/b19107-11>
- [19] H. Zare and M. Taghi Bahreyni Toossi, ser. *Digital Imaging and Computer Vision*. CRC Press, Sep 2015, ch. Early Detection of Melanoma in Dermoscopy of Skin Lesion Images by Computer Vision-Based System, pp. 345–384.
- [20] G. Fabbrocini, V. D. Vita, S. Cacciapuoti, G. D. Leo, C. Liguori, A. Paolillo, A. Pietrosanto, and P. Sommella, *Automatic Diagnosis of Melanoma Based on the 7-Point Checklist*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 71–107. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-39608-3\\_4](http://dx.doi.org/10.1007/978-3-642-39608-3_4)
- [21] F. Nachbar, W. Stolz, T. Merkle, A. B. Cogna, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The abcd rule of dermatoscopy," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 2016/04/22 XXXX. [Online]. Available: [http://dx.doi.org/10.1016/S0190-9622\(94\)70061-3](http://dx.doi.org/10.1016/S0190-9622(94)70061-3)
- [22] A. Sez, C. Serrano, and B. Acha, "Model-based classification methods of global patterns in dermoscopic images," *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1137–1147, May 2014.
- [23] C. Serrano and B. Acha, "Pattern analysis of dermoscopic images based on markov random fields," *Pattern Recognition*, vol. 42, no. 6, pp. 1052–1057, 2009. [Online]. Available: <http://dblp.uni-trier.de/db/journals/pr/pr42.html#SerranoA09>
- [24] M. Sadeghi, T. K. Lee, D. McLean, H. Lui, and M. S. Atkins, "Global pattern analysis and classification of dermoscopic images using textons," vol. 8314, 2012, pp. 83 144X–83 144X–6. [Online]. Available: <http://dx.doi.org/10.1117/12.911818>
- [25] A. G. Isasi, B. G. Zapirain, and A. M. Zorrilla, "Melanomas non-invasive diagnosis application based on the abcd rule and pattern recognition image processing algorithms," *Comp. in Bio. and Med.*, vol. 41, no. 9, pp. 742–755, 2011. [Online]. Available: <http://dblp.uni-trier.de/db/journals/cbm/cbm41.html#IsasiZZ11>
- [26] G. D. Leo, A. Paolillo, P. Sommella, G. Fabbrocini, and O. Rescigno, "A software tool for the diagnosis of melanomas," in *Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE*, May 2010, pp. 886–891.
- [27] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, April 2017.
- [28] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," vol. 542, 01 2017.
- [29] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Dec 2016, pp. 1–6.
- [30] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jan. 2015.
- [31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995. [Online]. Available: <https://doi.org/10.1007/BF00994018>
- [32] A. Marghoob, J. Malvehy, R. Braun, and A. Kopf, *An Atlas of Dermoscopy*, ser. Encyclopedia of Visual Medicine. CRC Press, 2004.
- [33] D. Pathak, P. Krähenbühl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in *ICCV*, 2015.
- [34] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterton, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256.
- [35] G. Argenziano, H. P. Soyer, and V. D. Giorgi, *Interactive Atlas of Dermoscopy*, 2002.
- [36] International Skin Imaging Collaboration: Melanoma Project, *ISIC Archive*, 2017. [Online]. Available: <https://isic-archive.com/>
- [37] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, "Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble," *CoRR*, vol. abs/1703.03108, 2017. [Online]. Available: <http://arxiv.org/abs/1703.03108>
- [38] I. González-Díaz, "Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions," *CoRR*, vol. abs/1703.01976, 2017. [Online]. Available: <http://arxiv.org/abs/1703.01976>
- [39] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. E. F. de Avila, and E. Valle, "RECOD titans at ISIC challenge 2017," *CoRR*, vol. abs/1703.04819, 2017. [Online]. Available: <http://arxiv.org/abs/1703.04819>
- [40] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," *CoRR*, vol. abs/1703.04197, 2017. [Online]. Available: <http://arxiv.org/abs/1703.04197>
- [41] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, "A novel multi-task deep learning model for skin lesion segmentation and classification," *CoRR*, vol. abs/1703.01025, 2017. [Online]. Available: <http://arxiv.org/abs/1703.01025>
- [42] P. Kisilev, E. Sason, E. Barkan, and S. Hashoul, *Medical Image Description Using Multi-task-loss CNN*. Cham: Springer International Publishing, 2016, pp. 121–129. [Online]. Available: [https://doi.org/10.1007/978-3-319-46976-8\\_13](https://doi.org/10.1007/978-3-319-46976-8_13)



**Iván González-Díaz** Iván González-Díaz received the Telecommunications Engineering degree from Universidad de Valladolid, Valladolid, Spain, in 1999, the M.Sc. and Ph.D. degree from Universidad Carlos III de Madrid, Madrid, Spain, in 2007 and 2011, respectively. After holding a postdoc position in the Laboratoire Bordelais de Recherche en Informatique at the University Bordeaux, he currently works as a Visiting Professor at the Signal Theory and Communications Department in Universidad Carlos III de Madrid. His primary research interests include object recognition, semantic image segmentation, scene understanding and content-based image and video retrieval systems. In these fields, he is co-author of several papers in prestigious international journals, two chapters in international books and a few papers in revised international conferences.