SUPORT MATERIAL: ON THE IMPACT OF SELF-SUPERVISED LEARNING IN SKIN CANCER DIAGNOSIS

Maria Rita Verdelho and Catarina Barata

Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal

1. FUSION OF THE ROTATION AND SIMCLR METHODS

This section gives a brief overview of the fusion of both SSL techniques.

We executed a set of experiments that combined both SSL techniques: Rotation and SimCLR. It is important to highlight that is also one of the novelties presented in this paper.

Both SSL techniques force the network to learn different tasks, which results in two models that might learn different information. However, the question that arises is: 'Is the information of both techniques complementary?'.

The goal of conducting these tests is to improve the global performance of both methods, assuming that each model carries different information about each skin lesion. In other words, the combination of both models can result in a more robust inference. To combine both Rotation and SimCLR techniques will be used the early and late fusion approaches, which will be addressed below. Both techniques differ at the level of fusion, early fusion concatenates the models in a feature level, while late fusion fuses the models in the classification scores levels [1].

1.0.1. Early Fusion

As the name indicates, early fusion combines the different methods in an earlier stage, which is in the feature space. This is known as feature level fusion and it consists of combining all the feature vectors into a single feature vector before sending them to the classifier. Therefore, this new model is trained to learn a correlation between the different features from the input models. In order to combine the set of models, the concatenation was used to jointly represent the different features. Figure 1 contains the pipeline used for the combination of the Rotation and the SimCLR technique using the early fusion.

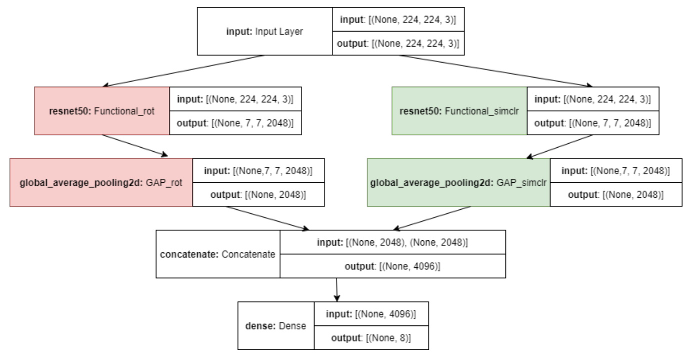

Fig. 1. Early Fusion Pipeline in red it is represented the rotation model and in green the SimCLR.

1.0.2. Late Fusion

Late fusion fuses the models in their final level, which is the classification scores level. Usually, this method uses fusion mechanisms that can consist of averaging, voting (which may result only with more than two models), or learned models.

In this paper, we decided to apply the mean value to the two score vectors and select the highest value that resulted in the combination of the final decision of both models. Figure 2 contains the pipeline used for the combination of the Rotation and the SimCLR technique using the late fusion.

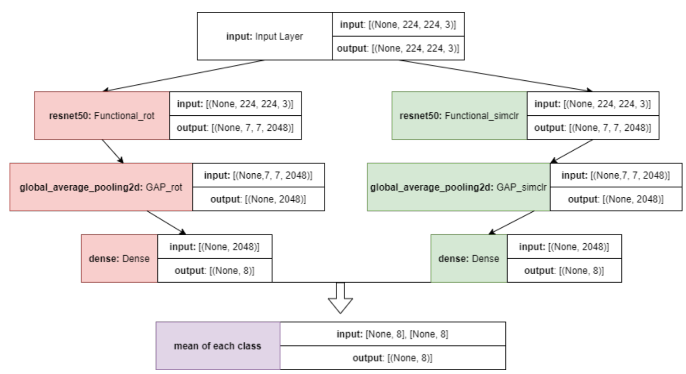

Fig. 2. Late Fusion pipeline in red it is represented the rotation model, in green the SimCLR and in purple the final score vector that results of mean of both score vectors.

2. QUANTITATIVE ANALYSIS PER PARTITION

This section presents the results obtained for the different partitions. Apart from proving consistency, this helps to prove that the fusion of different SSL features achieves the best results so that the studied two SSL methods learned complementary features.

Table 1. Application of the Monte Carlo Sampling with different initialization techniques using BACC: training the model from scratch or fine-tuning with ImageNet weights; application of two self-supervised learning (SSL) techniques -Rotation and SimCLR - and fusion of both techniques.

Initialization	Technique	Balanced Acc (%)						
		Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Median**	**standard deviation**
Baseline	Scratch	51,62	45,89	49,25	46,82	43,64	46,82	2,77
	Imagenet	70,57	75,09	71,48	73,08	70,21	71,48	1,82
Scratch + SSL	Rotation	54,23	56,83	54,92	55,21	53,36	54,92	1,15
	SimCLR	52,36	52,78	52,54	53,72	51,05	52,54	0,86
Imagenet + SSL	Rotation	71,47	71,77	71,10	71,95	71,32	71,47	0,31
	SimCLR	65,37	65,89	65,51	65,77	64,33	65,51	0,56
Imagenet + SSL	early fusion	73,78	73,90	73,53	74,18	73,57	73,78	0,24
	late fusion (mean)	59,00	59,09	57,09	59,69	53,85	59,00	2,19

Table 2. Application of the Monte Carlo Sampling with different initialization techniques using Precision: training the model from scratch or fine-tuning with ImageNet weights; application of two self-supervised learning (SSL) techniques -Rotation and SimCLR - and fusion of both techniques.

Initialization	Technique	Precision (%)						
		Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Median**	**standard deviation**
Baseline	Scratch	43,81	39,50	35,37	34,81	33,69	35,37	3,84
	Imagenet	68,09	63,29	65,14	65,73	59,81	65,14	2,78
Scratch + SSL	Rotation	40,54	43,79	38,38	39,40	41,16	40,54	1,84
	SimCLR	46,28	44,62	42,52	44,55	46,25	44,62	1,39
Imagenet + SSL	Rotation	62,81	60,87	61,67	62,73	62,37	62,37	0,74
	SimCLR	54,47	54,47	55,76	50,88	59,29	54,47	2,71
Imagenet + SSL	early fusion	70,64	66,87	65,52	68,41	70,82	68,41	2,07
	late fusion (mean)	48,82	51,19	51,78	48,25	49,36	49,36	1,39

Table 3. Application of the Monte Carlo Sampling with different initialization techniques using F1-score: training the model from scratch or fine-tuning with ImageNet weights; application of two self-supervised learning (SSL) techniques -Rotation and SimCLR - and fusion of both techniques.

Initialization	Technique	F1-score (%)						
		Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Median**	**standard deviation**
Baseline	Scratch	46,56	41,69	37,24	35,76	33,80	37,24	4,64
	Imagenet	68,44	67,52	67,93	68,92	64,02	67,93	1,75
Scratch + SSL	Rotation	43,29	46,81	40,71	41,94	43,19	43,19	2,04
	SimCLR	47,93	47,53	45,35	47,34	47,84	47,53	0,96
Imagenet + SSL	Rotation	66,12	65,02	65,55	66,36	65,70	65,70	0,47
	SimCLR	58,28	58,28	58,57	54,38	60,33	58,28	1,95
Imagenet + SSL	early fusion	72,02	67,30	65,95	70,99	72,10	70,99	2,61
	late fusion (mean)	51,80	52,35	52,64	51,69	49,89	51,80	0,96

3. QUANTITATIVE ANALYSIS PER CLASS OF LESION

Due to space constraints, the original document does not present the individual classification scores of the 8 lesion classes. Therefore in this section we present in table 5 the mean and standard deviation results of the diagonal values of the confusion matrices obtained for all 5 partitions. The diagonal of the matrix represents the SE by class, which would ideally be equal to 1, meaning that all samples in that class are correctly classified. Analyzing table 5 it is possible to verify the worst and best class of skin lesion classified by each method.

Table 4. Application of the Monte Carlo Sampling with different initialization techniques using SP: training the model from scratch or fine-tuning with ImageNet weights; application of two self-supervised learning (SSL) techniques -Rotation and SimCLR - and fusion of both techniques.

		SP (%)						
Initialization	**Technique**	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Median**	**standard deviation**
Baseline	Scratch	93,90	93,25	92,89	92,59	92,33	92,89	0,55
	Imagenet	96,14	96,03	96,04	96,14	95,81	96,04	0,12
Scratch + SSL	Rotation	93,39	93,54	93,04	93,30	93,49	93,39	0,18
	SimCLR	93,82	93,94	93,70	94,17	94,14	93,94	0,18
Imagenet + SSL	Rotation	95,81	95,75	95,77	95,80	95,67	95,77	0,05
	SimCLR	95,17	95,17	95,09	94,83	95,40	95,17	0,18
Imagenet + SSL	early fusion	96,58	95,83	95,80	96,40	96,57	96,40	0,36
	late fusion (mean)	94,51	94,29	94,18	94,50	93,96	94,29	0,21

Table 5. Mean and Standard Deviation values obtained in the 5-partitions for the normalized diagonal values of the confusion matrices.

	Classes							
Initialization	MEL	NV	BCC	AK	BKL	DF	VASC	SCC
Baseline	0,67 ± 0,02	0,80 ± 0,01	0,77 ± 0,02	0,61 ± 0,03	0,60 ± 0,02	**0,74** ± 0,06	0,83 ± 0,05	0,6 ± 0,09
Rotation	0,54 ± 0,04	**0,80 ± 0,00**	0,78 ± 0,04	0,63 ± 0,06	0,63 ± 0,04	0,72 ± 0,03	**0,84** ± 0,04	**0,62 ± 0,04**
SimCLR	0,59 ± 0,05	0,80 ± 0,03	0,78 ± 0,04	0,61 ± 0,05	0,64 ± 0,06	0,43 ± 0,09	0,79 ± 0,05	0,46 ± 0,08
Fusion	**0,72** ± 0,03	0,80 ± 0,03	**0,83** ± 0,05	**0,68** ± 0,03	**0,72** ± 0,01	0,71 ± 0,07	0,74 ± 0,07	0,60 ± 0,05

4. QUALITATIVE ANALYSIS

In this section we will present more results obtained using the Grad-CAM algorithm.

We present in figure 3 a set of five different lesions where the SimCLR model had difficulties in classifying the correct lesion. Analyzing this figure it is possible to verify that the SimCLR model had difficulties on detecting lesions which presented similar contrast to the skin. By looking at the three first examples (row 1, row 2 and row 3) it is visible that apart from having similar contrast to the skin, these images also present a higher contrast margin which may mislead the focus of the model.

Figure 4 presents a set of five different lesions where the Rotation model had difficulties in classifying the correct lesion. Analyzing this figure it is possible to verify that the Rotation model had difficulties on detecting lesions which also presented similar contrast to the skin (row 3, row 4 and row 5). By looking at the second example it is visible that the rotation had difficulties in detecting centered and symmetrical lesions, since each rotation of 90 degrees is similar, then the model does not learn useful information about this lesion. Finally, by looking at the first example it is possible to see that the rotation tends to find the non symmetrical areas of the images, however in this case it classified the lesion incorrectly. This could be explained by the fact that the lesion had a similar contrast to the background.

In figure 5 six different lesions where it is possible to see that each model looks at different parts of the input in order to

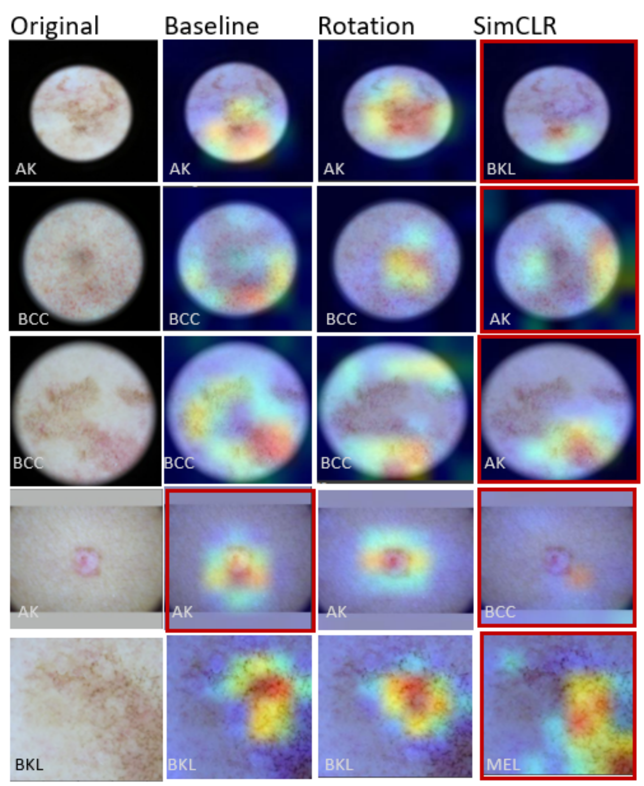

Fig. 3. Example of different lesion visualizations using the Grad-CAM algorithm (Baseline, Rotation and SimCLR). Showing limitations of the SimCLR model.

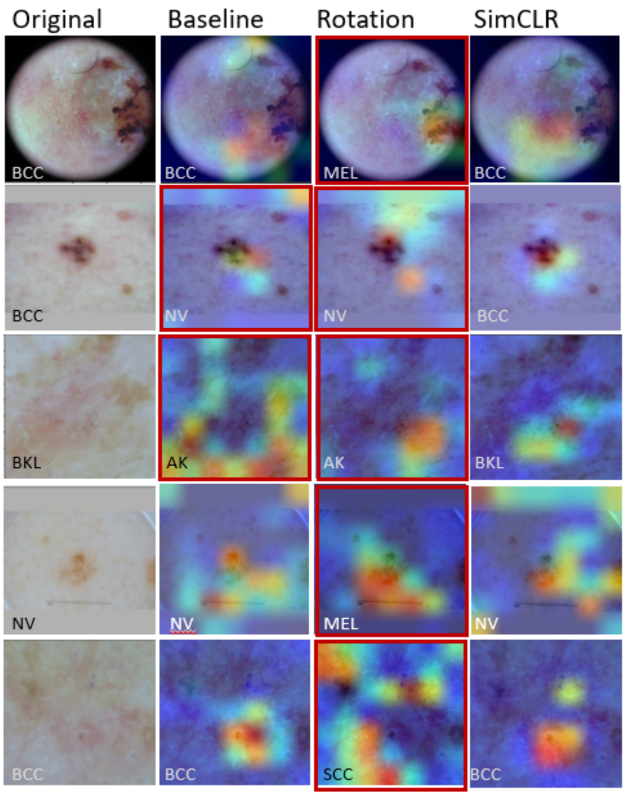

Fig. 4. Example of different lesion visualizations using the Grad-CAM algorithm (Baseline, Rotation and SimCLR).Showing limitations of the Rotation model.

make a decision. Apart from being able to detect each lesion, it is clear that each model looks at different features of the input. Meaning that they learn different information.

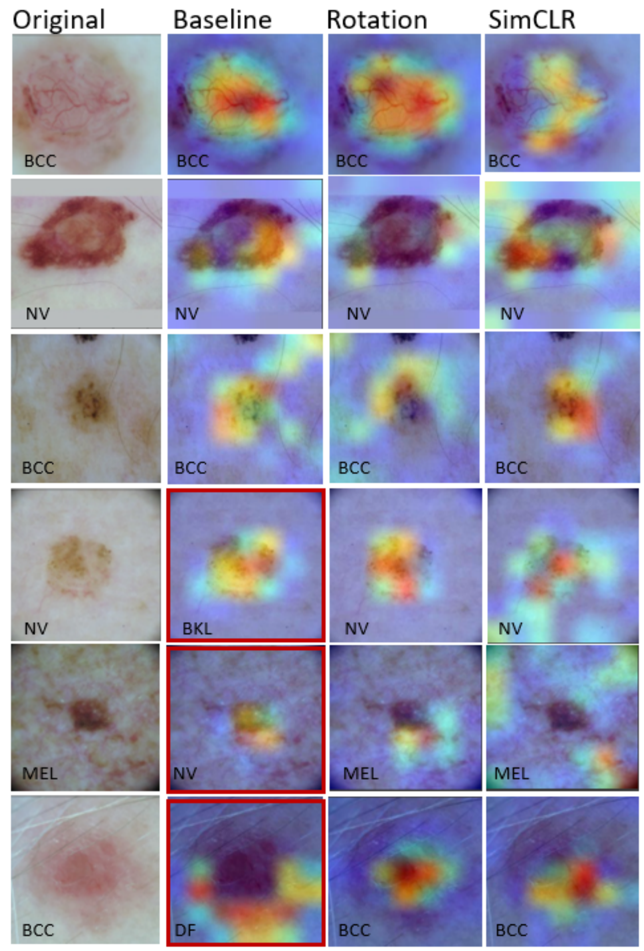

Fig. 5. Example of different lesion visualizations using the Grad-CAM algorithm (Baseline, Rotation and SimCLR).

5. REFERENCES

[1] Kuan Liu, Yanen Li, Ning Xu, and Prem Natarajan, "Learn to combine modalities in multimodal deep learning," 2018.