Avaliação de uma rede neural, do tipo *deep learning*, pré-treinada para reconhecer expressões faciais de uma base de imagens denominada AR\_Face

**Heraclito Amancio Pereira Junior e Saulo Caliman Gomes**

**Post-graduate Program in Informatics, Federal University of Espírito Santo, Espírito Santo, Brazil**

***Resumo :***  **Este trabalho relata a experiência no uso de uma rede neural artificial do tipo deep learning, pré-treinada, denominada de VGG Face, no reconhecimento de expressões faciais de uma base de imagens denominada AR\_Face. A rede neural faz parte do framework Caffe, criado pelo Berkeley Vision and Learning Center da Universidade de Berkley, e que foi pré-treinada por pesquisadores do Visual Geometry Group do Departamento de Ciências da Engenharia da Universidade Oxford no Reino Unido, para reconhecimento de faces. A AR\_Face contém mais de 4.000 imagens a cores correspondentes a 126 rostos de pessoas e foi criada por pesquisadores do Centro de Visão de Computador da University of Alabama at Birginham nos Estados Unidos. O objetivo desta pesquisa foi o de obter a melhor taxa de acurácia no treino da rede neural VGG Face utilizando-se uma base de imagens diferente daquela utilizada no pré-treinamento e modificações na estrutura original da rede pré-treinada.**

1. **Introdução**

Nos últimos anos o reconhecimento de faces tem se tornado uma área de pesquisa que tem despertado muito interesse da comunidade científica que atua na área de visão computacional. A visão computacional é um campo da ciência da computação que inclui métodos, técnicas e tecnologia para obtenção de informações simbólicas ou numéricas de imagens ou dados multidimensionais [1]. A expectativa geral é de que os avanços na pesquisa da visão computacional irão fornecer informações úteis para os neurocientistas e psicólogos sobre como cérebro humano funciona [2].

O reconhecimento de expressões faciais, em particular, tem sido visto como um grande potencial em diversas áreas tais como: jogos interativos, marketing, saúde e segurança. Apesar de ser uma tarefa trivial para o ser humano, reconhecer expressões não é ainda facilmente realizada pelos computadores.

As redes neurais artificiais são sistemas paralelos distribuídos compostos por unidades de processamento simples (neurônios artificiais) que calculam determinadas funções matemáticas. Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões e, na maioria dos modelos, essas conexões estão associadas a pesos, os quais armazenam o conhecimento adquirido pelo modelo e servem para ponderar a entrada recebida por cada neurônio da rede [3]. As redes neurais artiﬁciais têm sido aplicadas com sucesso nos mais diversos problemas da visão computacional, como reconhecimento de faces, reconhecimento de expressões faciais, reconhecimento de gestos, entre outros.

As redes neurais de aprendizagem profunda (*deep learning*) se distinguem das redes neurais mais comuns, de camada única escondida, por sua profundidade, isto é, pelo maior número de camadas através das quais os dados passam em num processo de reconhecimento de padrões. Esse tipo de redes é indicado para solução de problemas cujo espaço de entrada seja localmente estruturado de forma espacial, que é o caso do processamento de imagens e do processamento da linguagem [4].

Uma rede neural convolucional é um tipo de rede neural, inspirada nos processos biológicos, onde os neurônios são organizados para obter uma imagem de maior resolução a partir de uma com resolução inferior, tratando inclusive regiões de sobreposição no campo visual, sendo, por isso, muito indicada para reconhecimento de imagens e vídeo [5].

Caffe é um framework para redes neurais artificiais do tipo deep learning criado por pesquisadores do Berkeley Vision and Learning Center da Universidade de Berkley, que reduz o trabalho de seu usuário através do uso de uma linguagem simples de parametrização. Ele se destaca, em comparação com outros frameworks deste tipo, pela grande abrangência de aplicações e pela desempenho superior quando utilizado com as bibliotecas OpenCV, BLAS e a CUDA da NVIDIA [6].

Este trabalho relata a experiência no uso de uma rede neural artificial do tipo deep learning, pré-treinada, denominada de VGG Face, que faz parte do framework Caffe, no reconhecimento de expressões faciais de uma base de imagens denominada AR\_Face. A VGG Face foi pré-treinada por pesquisadores do Visual Geometry Group do Departamento de Ciências da Engenharia da Universidade Oxford no Reino Unido [6]. A AR\_Face contém mais de 4.000 imagens a cores correspondentes a 126 rostos de pessoas e foi criada por pesquisadores do Centro de Visão Computacional da University of Alabama at Birginham nos Estados Unidos (CVC). As imagens apresentam rostos com vista frontal e diferentes expressões faciais, condições de iluminação e oclusões (óculos de sol e lenço). As fotos foram tiradas na CVC, sob condições estritamente controladas, tendo os participantes sido submetidos ao uso de determinados acessórios de vestuário (lenços, echarpes, óculos, etc.), maquiagem, estilo de cabelo, etc.. Cada pessoa participou em duas sessões, separadas por duas semanas. As mesmas fotos foram tiradas em ambas as sessões [7].

O objetivo da pesquisa foi o de obter uma melhor taxa de acurácia no treino da rede neural VGG Face utilizando-se uma base de imagens diferente daquela utilizada no pré-treinamento e modificações na estrutura original da rede pré-treinada.

**II. Trabalhos Relacionados**

A procura por uma solução que propicie um reconhecimento de faces com a máxima confiabilidade, baseada em taxas médias de acurácia no reconhecimento, tem produzido, há algum tempo, uma série de pesquisas em diversas partes do mundo. Dentre as técnicas empregadas para reconhecimento de expressões faciais tem-se destacado o uso de redes neurais artificiais.

Através de consultas ao Portal da Capes e ao Google Scholar pôde-se encontrar uma série de trabalhos publicados sobre o uso de redes neurais artificiais para reconhecimento de expressões faciais, dentre os quais, discorre-se a seguir sobre aqueles mais citados ou cujas pesquisas mais se aproximam deste trabalho.

As pesquisas com o uso de redes neurais artificiais já ocorrem há algum tempo. Em 1991, Kobayashi e Hara [8] desenvolveram um sistema para reconhecimento de emoções humanas utilizando as redes neurais artificiais para classificação de expressões faciais. Os autores utilizaram as seis expressões básicas da emoção como categorias para a classificação das redes neurais. A rede neural foi treinada utilizando 90 imagens com as seis expressões faciais básicas de 15 indivíduos sendo obtida uma taxa de reconhecimento média próxima de 80 por cento.

Em 1996, Padgett e Cottrell [9] utilizaram uma rede neural com uma camada escondida para classificação do tipo sigmóide não-linear, com 7 saídas, sendo 6 saídas para cada uma das emoções básicas e uma para as expressões neutras. Foram utilizadas diversas imagens de 12 indivíduos e a taxa média de reconhecimento foi de 86 por cento.

A pesquisa relatada em [10] demonstrou que o uso de uma rede neural convolucional poderia ser mais adequada nas situações para reconhecimento de faces e, por consequência também para o reconhecimento de expressões faciais. Nela foi apresentada uma técnica baseada em uma arquitetura de rede neural que, sem localizar nenhum ponto específico da face, conseguia alinhar precisamente as imagens faciais.

As pesquisas nessa área continuaram a correr e uma visão geral dos trabalhos realizados entre 2002 e 2012 pode ser vista em [11].

Dentre os trabalhos encontrados mais recentemente que mais se aproximam das condições presentes nesta pesquisa, pode-se citar os artigos publicados pelos pesquisadores da Berkeley Vision and Learning Center da Universidade de Berkley e do Visual Geometry Group do Departamento de Ciências da Engenharia da Universidade Oxford no Reino Unido. Em 2010, no evento [ImageNet Large Scale Visual Recognition Competition 2010](http://image-net.org/challenges/LSVRC/2010/), os pesquisadores da Berkley treinaram uma rede neural convolucional do tipo deep learning, para classificar 1,3 milhão imagens de alta resolução e obtiveram resultados consideravelmente melhores do que os resultados anteriores do estado-da-arte. Usaram uma rede neural com 60 milhões de neurónios em cinco camadas convolucionais. Para tornarem o treinamento mais rápido, utilizaram uma implementação com a biblioteca CUDA para realizar as operações da rede na GPU (Graphics Processor Unit ) [5]. Ainda dos pesquisadores da Berkley pode-se destacar a apresentação do framework Caffe de redes neurais do tipo deep learning. Esse framework inova por disponibilizar modelos de redes neurais pré-treinadas que podem ser facilmente reconfiguradas para adequá-las às necessidades dos pesquisadores [12].

Em [6] os pesquisadores de Oxford apresentam duas contribuições importantes: i) como montar um grande banco de imagens, com mais de 2,6 milhões imagens de mais de 2600 pessoas, utilizando filtros computacionais para seleção das imagens e ações manuais para filtrar e remover imagens indesejadas, e ii) utilização de uma rede neural artificial para treinar e testar o reconhecimento de faces a partir do banco de imagens gerados com resultados expressivos (acurácias entre 92 e 96%), em relação a experiências exitosas anteriores. A implementação da rede neural se baseou no Tool Box MATLAB MatConvNet [13] combinado com as bibliotecas da NVIDIA CuDNN, que é uma extensão da biblioteca CUDA, para aceleração do treino. Foi utilizada uma configuração computacional robusta com quatro placas GPU NVIDIA Titan Black com 6 GB de memória interna cada uma.

**III. Metodologia**

A metodologia para a realização deste trabalho foi composta das seguintes etapas- atividades:

**Etapa 1:** Estudos e Seleção de Rede Neural Pré-Treinada - ***Atividades:***

1.1 - Levantamento de artigos sobre uso de redes neurais para reconhecimento de faces e de expressões faciais;

1.2 - Estudo destes artigos;

1.3 - Estudo do Tutorial na web sobre o framework Caffe e sobre os procedimentos de como obter e instalar uma rede Caffe;

1.4- Escolha de uma rede neural pré-treinada dentro da Zoo Model. A Zoo Model é uma página web contendo diversos modelos de redes neurais pré-treinadas baseados no Caffe, configuradas com diversas arquiteturas, que foram aplicadas a diversos tipos de problemas. Naquela página, existem informações sobre como obter os arquivos com binários das redes pré-treinadas, arquivos modelos de configuração e ferramentas para fazer donwload e upload desses arquivos.

**Etapa 2:** Preparação do Ambiente para o Experimento ***Atividades:***

2.1 - Preparação do ambiente computacional para instalação da rede Caffe e dos arquivos referentes à rede pré-treianada escolhida;

2.2 - Instalação das bibliotecas: CUDA, Open CV e BLAST;

2.3- Instalação da rede Caffe;

2.4 - Download dos arquivos da rede pré-treinada;

2.5 - Download de uma base de imagens menor para os primeiros testes com a rede escolhida;

2.6 - Preparação da rede através da configuração dos arquivos de configuração da rede e de seus parâmetros de processamento:“Net.Prototxt” e “Solver.Prototxt;

2.7 - Preparação dos arquivos de configuração das visões das imagens de treino e das imagens de teste;

2.9 - Download da base de imagens AR\_Face;

2.10 - Análise da base de imagens AR\_Face e depuração de suas imagens para o propósito de reconhecimento de expressões.

**Etapa 3:** Execução dos Testes, Análises e Ajustes **-**

***Atividades:***

3.1 - Execução dos primeiros testes da rede neural para a validação do ambiente e da instalação do Caffe;

3.2 - Ajuste nas configurações dos arquivos de configuração “Net.Prototxt” e “Solver.Prototxt e dos arquivos de configuração das visões das imagens de treino e de teste, agora para a base de imagens AR-Face;

3.3 - Processamento da rede pré-treinada escolhida com a base de imagens AR\_Face;

3.4 - Análise dos Resultados (principalmente a acurácia obtida);

3.5 - Novos ajustes nos arquivos de configuração e visões da rede para treino e testes.

**Etapa 4:** Análises Finais e Elaboração do Relatório**:**

***Atividades:***

4.1 - Análise dos resultados finais;

4.2 - Escrita sobre o experimento e seus resultados.

A Figura 1 a seguir mostra um diagrama com a configuração para processamento da rede.

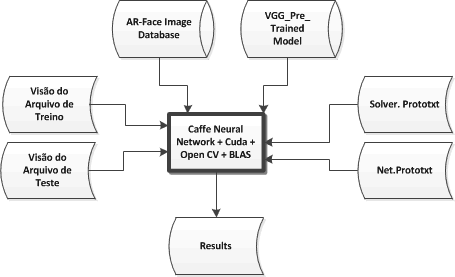


Figura 1 - Esquema de processamento da rede neural Caffe para reconhecimento de expressões faciais

**IV. Desenvolvimento do Experimento**

**Etapa 1:** Estudos e Seleção de Rede Neural Pré-Treinada:

Foram executados os levantamentos de artigos e livros o sobre redes neurais, redes neurais do tipo deep learning, redes neurais convolucionais e sobre o uso de redes neurais para reconhecimento de faces e de expressões faciais, resultando nas referencias apontadas na parte final deste relatório.

Foram lidos os artigos e o tutorial na web sobre o framework Caffe e os procedimentos de como se obter e instalar uma rede Caffe.

Pesquisou-se no site do framework da Caffe (<http://caffe.berkeleyvision.org/>) para identificar modelos de redes pré-treinadas sobre reconhecimento de faces que pudessem ser utilizadas para os experimentos previstos para a pesquisa. No site do framework Caffe, além de outras informações, encontrou-se um link onde estavam as referências das redes pré-treinadas: Model Zoo Models (https://github.com /BVLC/caffe/ wiki/Model-Zoo).

Analisando as informações daquelas redes pré-treinadas, optou-se pela rede neural pré-treinada VGG Face CNN (Convulotional Neural Network) originada de um trabalho dos pesquisadores do Visual Geometry Group do Departamento de Ciências da Engenharia da Universidade Oxford no Reino Unido [6]. O trabalho do VGG Group utilizou um banco de imagens com 2,6 milhões de imagens de celebridades coletadas da web e teve dois objetivos: propor um procedimento para criar um grande banco de imagens de faces humanas e investigar várias arquiteturas CNN para identificação de faces humanas. O modelo do VGG Group utilizou o framework Caffe e se baseou no modelo pré-treinado MatConvNet [13].

A escolha da rede pré-treinada VGG teve foi devida ao fato dela já ter sido pré-treinada para reconhecimento de faces, ter obtido ótimos resultados de acurácia de reconhecimento e ser uma originada de uma pesquisa recente (de 2016). A Fiigura 2 mostra a configuração original da rede pré-treinada VGG Face CNN.

**Etapa 2:** Preparação do Ambiente para o Experimento

Inicialmente optou-se pela instalação de da Rede Caffe para ambiente Windows, pois o computador disponível e com recursos de hardware suficientes para atender os requisitos para processamento de uma rede neural Caffe do tipo deep learning já tinha este sistema operacional instalado.

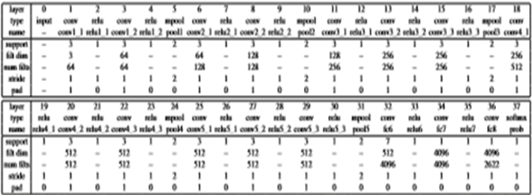


Figura 2 – Configuração original da rede neural pré-treinada VGG Face CNN

Utilizou-se então o tutorial “Build Caffe in Windows with Visual Studio 2013 + CUDA 6.5 + OpenCV 2.4.9 publicado em janeiro de 2015 por Neil Z. Shao (<https://initialneil.wordpress.com> /2015/01/11/build-caffe-in-windows-with-visual-studio-2013-cuda-6-5-opencv-2-4-9/).

O ambiente começou então a ser preparado, em 01/05/2016, em um computador I7, com 8GB de Memória RAM e uma Placa de Vídeo Nvidia GTX 970 de 4 GB.

Foi utilizada a versão do CUDA 6.5 com driver nativo da NVIDIA 340.29.

Foi instalada a biblioteca de identificação de imagens, OPENCV 2.4.10 e sua instalação ocorreu conforme estava prevista.

Utilizou-se inicialmente, para os primeiros testes, um conjunto de imagens denominado *YALE-Expressions* contendo 15 conjunto de imagens, cada conjunto referente a uma pessoa e seis expressões faciais básicas delas.

A seguir, foi feita a preparação do arquivo de configuração do processamento da rede “Solver.Prototxt “, tendo sido ajustados os parâmetros:

* **test\_iter = 1**  (inicialmente para 1 pois o arquivo de testes inicial conteria um conjunto de 6 imagens de uma das pessoas e previa-se um que o parâmetro *batch size* para a configuração da rede teria valor igual a 6. Segundo as orientações dos pesquisadores da Caffe a multiplicação do test\_iter pelo batch\_size deveria ser igual ao número de imagens testadas);
* **test\_interval =** 100 (cada teste ocorreria após cada 100 iterações de treino);
* **base\_lr = 0.001** (segundo as orientações dos pesquisadores da Caffe esta taxa de aprendizado poderia começar entre 0.01 e 0.001 e posteriormente ser ajustada de acordo com a percepção da a evolução da aprendizagem);
* **lr\_policy = "inv", gamma = 0.001, power = 0.75** (estes parâmetros que identificam a queda da taxa de aprendizagem foram mantidos de acordo com a configuração original pois entendeu-se que a taxa de aprendizado poderia ser reduzida à medida que os treinos da rede fossem ocorrendo, até que a perda atingisse um platô e começaria a se repetir indefinidamente, conforme mostra a Figura 3. Estes parametros foram usados por Krizhevsky et al. Em [5].

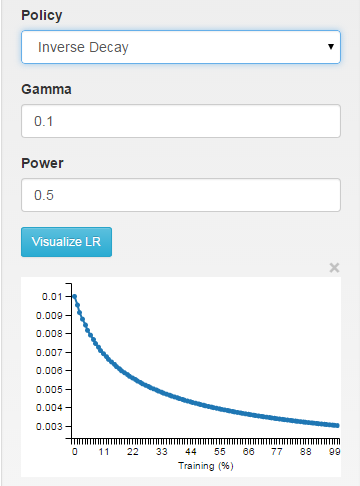
. 

Figura 3 – Queda da taxa de aprendizado da rede conforme o tempo de treino.

* Os parâmetros número de interações a partir das quais os resultados seriam mostrados no vídeo(**display**), número máximo de interações para a rede (**max\_iter** ), período de interações para gravação das posições da rede treinada até um determinado momento do treino (**snapshot**) e o tipo de processamento com uso ou não da placa gráfica do computador (**solver\_mode** ) foram ajustados conforme aa seguir:

display: 100

max\_iter: 10000

snapshot: 1000

snapshot\_prefix: "snapshots/"

solver\_mode: GPU

Em relação à configuração da rede original VGG (arquivo Net.Prototxt) , tomou-se as seguintes providencias:

* Ajustes nas camadas de entrada de dados: (para colocar o nome dos arquivos de configuração das imagens para treino e teste, ajuste no parâmetro batch\_size para o número de imagens que seriam carregadas na memória, tomando o cuidado para não carregar muitas imagens ao mesmo tempo para não ultrapassar a memória RAM disponível). Colocando-se um **batch\_size** pequeno o processamento de treino se torna mais rápido porém tem-se que acompanhá-lo para se evitar o problema de geração de muita oscilação na acurácia, conforme recomendação encontrada em fóruns sobre a rede Caffe.
* Ajuste inicial no parâmetros **blobs\_lr** (para valores baixos) nas camadas de convolução (camadas que tratam da qualidade das imagens) para que elas ficassem um pouco “congeladas” e propiciassem então uma otimização rápida das últimas camadas de classificação conforme recomendação encontrada em fóruns sobre a rede Caffe);
* Ajustes nas camadas de classificação (**INNER- PRODUCT)** para que elas tivessem melhores taxas de aprendizagem e fornecessem como saída as 6 expressões faciais que a base de imagens iniciais continham.

**Etapa 3:** Execução dos Testes, Análises e Ajustes:

(Todas as telas com os resultados dos testes estão no Anexo 01).

- A execução dos testes se iniciou ainda no dia 01/05/2016, utilizando-se a rede Caffe no ambiente Windows. Após as 10.000 iterações, programadas no arquivo de configuração Solver. Prototxt, terminaram sem que se conseguisse gravar as telas dos resultados do processamento de forma a verificar qual foi a acurácia obtida.

- No dia 02/05/2016foi executado um processamento da rede que, inicialmente apresentou uma animadora acurácia de treino e de teste de 20%, porém a rede terminou de processar as 10.000 iterações, que estavam programadas, sem que tais acurácias evoluíssem.

- No dia 07/05/2016, entendendo-se que o número de iterações máximas podia ser um limite para a melhoria da acurácia, aumentou-se seguidamente o número máximo de iterações para três patamares superiores : 1º - de 10.000 para 100.000, 2º - de 100.000 para 150.000 e 3º - de 150.000 para 200.000. Entretanto, após cada mudança, a acurácia manteve-se em 20%.

- No dia 12/05/2016, supôs-se que a política para a taxa de aprendizagem do tipo INV(**lr\_policy**) poderia estar causando o rápido crescimento da acurácia para 20% mas depois decaindo até uma baixíssima taxa de aprendizagem, o que poderia estar impedindo o crescimento da acurácia a partir dos 20%. Mudou-se então a política para a taxa de aprendizagem para “**FIXED**”. Após 9.400 iterações da rede, a acurácia máxima de teste e a máxima de treino obtida diminuiu pra 16,66%, permanecendo o comportamento anterior de chegar rápido ao valor máximo e estagnar nele.

- No dia13/05/2016, cogitou-se que a composição dos arquivos de teste e treino (que estavam respectivamente com 10% e 90% do conjunto de imagens do base de imagens que se estava utilizando ) poderia estar impactando o crescimento da acurácia. Ao se analisar o arquivo de teste teve-se a impressão de que a pessoa escolhida para ter as suas seis expressões faciais testadas seria bem diferente das outras pessoas. Elaborou-se novas composições de arquivos de teste e treino foram processadas (30% de teste e 70% de treino e, depois, 50% de teste e 50% de treino) sem que se obtivesse melhoria da acurácia.

- No dia 14/05/2016, após a reanálise das configurações da rede neural sem encontrar outros motivos pelos insucessos até o momento, supôs-se que a instalação Windows da rede Caffe não era confiável e decidiu-se por abandoná-la e instalar a sua versão para o Ubuntu 12.04, no mesmo computador que se estava usando até o momento.

- Entre os dias 16 e 19/05/2016, foi feita a instalação da versão da Caffe para o Ubuntu 12.04, no computador que se estava usando anteriormente e, durante essa instalação surgiu uma série de problemas. A título de ilustração dos problemas ocorridos descreve-se a seguir alguns deles.

* Não se conseguia baixar a biblioteca *ippicv\_linux20151201* diretamente pelo procedimento de instalação e a solução foi de baixá-la diretamente do site;
* Foi encontrado erro no arquivo hdf5.cpp.o e, após a verificação de existência através do comando *dpkg -l | grep hdf5*, foi identificada a sua versão (1.8.4) e a solução foi a de retirar o parâmetro de compilação do HDF;
* Na instalação do OpenCV-master, obtido a partir do *branch master do github*, ao se executar o Cmake, foi apresentado o mesmo erro da falta da biblioteca *ippicv\_linux20151201* e a solução adotada foi a mesma;
* Ao se executar o “make” não foi encontrada a classe *reflection no protocol buffer do google*. Verificou-se que ela existia no *protocol buffer 3.0.0 betaV3* e a solução foi então atualizar a versão do *protocol buffer*;
* Na compilação do *protobuffer* apareceu a necessidade do *gmock-1.7.0* e a solução foi baixa-lo.
* Após a instalação do *libtool* para a build correta foi necessário o *autoconf* para gerar *o ./configure;*
* A execução do *protobuffer 3.0.0* falhou e, procurando nos fóruns, foi encontrada uma versão 2.5.0 recomendada como solução para o problema.

- Após a conclusão da instalação da rede neural VGG no ambiente Ubuntu, no computador que estava sendo usado anteriormente, aconteceram vários problemas de “out of memory” e várias tentativas de reconfiguração da rede (arquivo Net.Prototxt) foram feitas, tais como: a diminuição do parâmetro **batch\_size** e modificação do tamanho do kernel das camadas de convolução, etc.) até que o processamento ficou normalizado. Com a superação do problema de “out of memory”, alguns processamentos foram feitos, entretanto o tempo de processamento ficou muito alto, inviabilizando o uso daquela instalação. Decidiu-se então pela utilização de computadores do laboratório LCAD que possuíam mais recursos de hardware.

- No dia 20/05/2016, foi feito o primeiro processamento da rede Caffe VGG utilizando-se um computador do laboratório LCAD, mais potente em termos de recursos de hardware, e com sistema operacional Ubuntu 12.04 LTS e bibliotecas Open CV 2.4.9 e Cuda 7. O computador usado tem a seguinte configuração: Processador Intel Xeon, Memória RAM de 12 GB e duas placas gráficas de 2 GB cada uma. Neste processamento utilizou-se ainda a mesma configuração arquivos Solver.Prototxt e Net.Prototxt, que já estava sendo usada no computador anterior e, também, com o banco de imagens contendo apenas 15 conjuntos de 6 imagens de uma mesma pessoa, cada uma imagem com uma expressão facial diferente. O resultado melhorou substancialmente tendo se obtido uma acurácia de 60% para testes e, várias vezes acurácia de 100% para treino.

- No dia 25/05/2016, antes de novos processamentos de treino da rede, fez-se uma depuração no base de imagens AR\_Face, que era composto por 76 conjuntos de imagens de homens e 60 conjuntos de imagens de mulheres, cada conjunto com 26 fotos da mesma pessoa em situações diferentes, eliminando-se aquelas fotos que não permitiam a correta identificação das faces (fotos com óculos, fotos com echarpes, fotos com brilho, etc.). Após essa depuração, os conjuntos ficaram compostos por 9 fotos por pessoa. Foram então criados os arquivos de teste com 10% das imagens disponíveis e de treino com 90% das imagens disponíveis. Utilizou-se a mesma configuração do arquivo Solver.Prototxt e no arquivo da configuração da rede (Net.Prototxt) promoveu-se as seguintes alterações:

* Nas camadas de entrada de dados, o batch-size foi alterado de 6 para 9 (pretendia-se se carregar apenas nove imagem de cada vez na memória para não se ter problema de “*out of memory*” e ter um processamento mais rápido). Também os parâmetros de altura e largura das imagens e de nomes dos arquivos de treino e teste foram ajustados;
* Nas camadas de convolução, foram alterados os parâmetros **blobs\_lr** passando-os para 0.1 e 0.2, visando com que essas camadas passassem a contribuir com os ajustes na qualidades das imagens.
* Nas camadas de classificação foi alterado o output para 9 (nove tipos de expressões possíveis).

Feito o processamento da rede VGG Cafee no computador do laboratório LCAD, foi obtido uma acurácia de 52,55% para testes e, várias vezes acurácia de 100% para treino. Os testes foram repetidos com algumas modificações em outros parâmetros de configuração da rede (Net.Prototxt) tais como: o parâmetro stride ( intervalos de aplicação dos filtros ás imagens de entrada) e pad (número de pixel adicionados em cada lado das imagens de entrada) das camadas de convolução visando melhorar o tratamento convolucional das imagens, sem que houvesse melhoria da acurácia.

- Entre os dias 01 e 02/06/2016 :

Alguns processamentos foram feitos tendo sido realizados os seguintes ajustes entre eles:

* Verificou-se que na primeira depuração do base AR\_Face a escolha de 9(nove) imagens como representativas de nove expressões foi equivocada pois deveriam ser apenas 4 pois uma das 9 se diferenciava apenas por ter brilho e as outras oito, duas a duas eram muito parecidas. Então se optou para eliminar uma imagem de cada conjunto e classificar duas as duas como representativas da mesma expressão. A base de imagens de treino foi então acertada.
* Optou-se também por crescer o número de imagens dos arquivos de teste para 25% da base e o arquivo de treino passou para 75% do base.
* Uma série de outros ajustes forma feitos nos arquivos de configuração do processamento (Solver. Prototxt) e no de configuração da rede (Net.Prototxt), a saber:

**Solver. Prototxt**:

* **test\_iter = passou-se de 1 para 55** (pois com 4 expressões o batch size deveria passar de 9 para 4 e, segundo as orientações dos pesquisadores da Caffe a multiplicação do test\_iter pelo batch\_size deveria ser igual ao número de imagens testadas);
* **lr\_policy = passou de "inv" para “fixed”** para que não houvesso o rápido decaimento da taxa de aprendizagem.

**Net.Prototxt:**

* nas camadas de entrada de dados, o batch-size foi modificado para 4 ;
* nas camadas de classificação foi alterado o output para 4 (nove tipos de expressões possíveis).

Feito então feito alguns processamentos no computador do laboratório LCAD, contemplando tais alterações, obtendo-se um resultado de **acurácia de**  **83,66%** para testes e, várias vezes acurácia de 100% para treino.

**VI. RESULTS AND CONCLUSIONS**

A utilização de uma rede pré-treinada através do framework Caffe é relativamente simples porém as maiores dificuldades encontradas foram :

1. Encontrar material mais didático sobre o assunto. Existem alguns materiais na WEB, inclusive o próprio tutorial da Caffe, porém encontramos algumas dificuldades para entender : melhor o funcionamento das camadas da rede e a as configurações dos parâmetros dos arquivos “Solver.Prototxt” e “Net.Prototxt”. Os próprios artigos encontrados sobre a utilização das redes Caffe não esclareciam bem esses assuntos;
2. A instalação da rede Caffe com as necessárias bibliotecas. Os tutoriais encontrados não são abrangentes e detalhados suficientes para garantir que seguindo-os chega-se facilmente a instalação da rede Caffe;
3. Outro problema encontrado foi que após a instalação é concluída em um determinado computador, não há como se garantir que o processamento da rede irá ocorrer normalmente. Não se consegue previamente estimar quais recursos, principalmente de memória, vão ser necessários a uma determinada configuração de rede. Apesar de se poder fazer ajustes nos parâmetros nos arquivos de configuração, uma vez que você conheça o significado de cada parâmetro, este processo ainda parece ser de mais de tentativas-e-erros.

Consideramos o resultado obtido de 84% de acurácia nos testes como um bom resultado considerando os seguintes motivos:

1. O arquivo final usado continha uma quantidade pequena de imagens comparadas aos trabalhos correlatos em que os resultados foram superiores. Por exemplo o trabalho relatado em [6], que obteve acurácias de 87% a 96%, utilizou uma base de 2,6 milhões de imagens enquanto a nossa base era de centenas de imagens. Nossa base ficou pequena porque o database especificado pelo enunciado do trabalho era o AR-Face - continha apenas 3500 imagens - e ficou menor ainda após o tratamento manual ao que tivemos que submetê-lo, para eliminar imagens que apresentavam obstáculos ao reconhecimento de expressões faciais, alinhamentos nas imagens e classificação das imagens de acordo com tipo de expressões faciais;
2. Os trabalhos correlatos que utilizaram conjuntos de imagens equivalentes ao nosso em tamanho [8, 9, 10] obtiveram acurácias parecidas com as nossas;
3. As tentativas que fizemos em aumentar o número de camadas de classificação para apurar a acurácia acabaram por gerar processamentos sem sucesso devido a problemas de “out fo memory”. Acreditamos que, com um computador com placas de vídeo mais poderosas poderemos obter melhores resultados na acurácia utilizando mais acamadas de classificação.

**REFERENCES**

[1] - Lisetti, Christine L., and David E. Rumelhart. "Facial Expression Recognition Using a Neural Network." *FLAIRS Conference*. 1998.

[2] - Face Recognition Homepage,<http://www.face-rec.org/general-info/> acessado em 14-05-2016

[3] - de Pádua Braga, Antônio, André Carlos Ponce de Leon Ferreira, and Teresa Bernarda Ludermir. *Redes neurais artificiais: teoria e aplicações*. LTC Editora, 2007.

[4] - Kishore, Anurag, Stuti Jindal, and Sanjay Singh. "Designing Deep Learning Neural Networks using Caffe." (2015).

[5] - Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[6] - Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." *Proceedings of the British Machine Vision* 1.3 (2015): 6.

[7] - A.M. Martinez and R. Benavente. The AR Face Base. CVC Technical Report #24, June 1998.

[8] - Kobayashi, Hiroshi, and Fumio Hara. "The recognition of basic facial expressions by neural network." *Neural Networks, 1991. 1991 IEEE International Joint Conference on*. IEEE, 1991.

[9] - C. Padgett and G.W. Cottrell, ªRepresenting Face Images for Emotion Classification,º Proc. Conf. Advances in Neural Information Processing Systems, pp. 894-900, 1996.

[10] - Lawrence, Steve, et al. "Face recognition: A convolutional neural-network approach." *Neural Networks, IEEE Transactions on* 8.1 (1997): 98-113.

[11] - C.-D. Caleanu, “Face expression recognition: A brief overview of the last decade,” in 2013 IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2013, pp. 157–161.

[12] - Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014.

[13] A. Vedaldi and K. Lenc. Matconvnet – convolutional neural networks for matlab. CoRR,

abs/1412.4564, 2014.