



Figura 5: Exemplo do aumento artificial de dados em uma imagem original, mostrada no painel (a). Os painéis (b), (c), (d), (e), (f) e (g) contêm os resultados da equação (3) substituindo M por diferentes combinações das transformações da equação (2). Em (b) $M = V$, em (c) $M = H$, em (d) $M = R(30^\circ)$, em (e) $M = VR(30^\circ)$, em (f) $M = HR(30^\circ)$ e em (g) $M = HVR(30^\circ)$.

3.1.1. Agrupamento das bandas para confecção das imagens RGB

Como as imagens do S-PLUS foram obtidas em 12 bandas fotométricas (listadas em [9]), para representá-las no sistema de cor RGB fizemos o seguinte mapeamento: em R colocamos as 4 bandas vermelhas r_SDSS, i_SDSS, J0861 e z, em G as bandas g_SDSS J0515 e J0660 e em B as cinco bandas mais azuis u_JAVA, J0373, J0395, J0410 e J0430 (as características destes filtros são dadas na Tabela 1 de [9]). Na combinação de bandas em cada canal, foi feita uma soma simples dos valores dos píxeis. Depois de reduzidas a três bandas, as imagens são usadas como entrada do programa Trilogy[11]⁵.

3.1.2. ImageNet

Como já mostrado em trabalhos anteriores, a inicialização dos pesos provenientes de uma rede pré-treinada usando a base de dados *ImageNet*⁶ traz uma grande melhoria na precisão dos resultados da classificação. Essa base de dados possui milhões de imagens de objetos do cotidiano e já foi utilizada especificamente para a classificação de objetos astronômicos (veja por exemplo [12]), com excelentes resultados.

O uso deste dataset para pré-treinamento respeitou o pré-processamento utilizado originalmente pelos autores de cada rede, este procedimento este foi crucial para garantir um fit competitivo, isto é, no *benchmarking* original destas redes para os dados da *ImageNet*. Para a rede VGG16 (Seção 3.3), a ordem das bandas foi trocada de RGB para BGR, e cada banda foi centrada em zero em relação à *ImageNet*, sem escalonamento, ou seja, os píxeis de cada banda tiveram o valor da média da respectiva banda *ImageNet* subtraído. Para a rede InceptionResNetV2 (Seção 3.4), os píxeis de entrada foram escalonados entre -1 e 1 em relação a amostra de treino. Para a rede EfficientNet (Seção 3.5), os píxeis de entrada foram escalonados entre 0 e 1 em relação à amostra de treino. E, para a rede DenseNet (Seção 3.6), os píxeis de entrada foram escalonados entre 0 e 1 e cada banda foi padronizada em relação à *ImageNet*, isto é, os píxeis de cada banda tiveram o

valor da média subtraído e o resultado foi dividido pelo desvio padrão da distribuição da respectiva banda da *ImageNet*.

3.2. Aumento artificial de dados

Aumento artificial de dados [13] é a aplicação de transformações afins nas imagens do conjunto de treinamento, por exemplo rotação, reflexão, translação e mudança de escala. As matrizes da equação (2) definem as transformações usadas.

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad V = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

onde $R(\theta)$ é a transformação rotação por um ângulo θ , H é a transformação reflexão horizontal e V é a transformação reflexão vertical.

Seja M a matriz das transformações combinadas, (x, y) a coordenada do píxel da imagem original e (x^*, y^*) a coordenada transformada do píxel, as transformações nas imagens são feitas remapeando as coordenadas dos píxeis originais aplicando uma combinação das matrizes da equação (2) em cada píxel da imagem original usando a equação (3), onde (t_x, t_y) é a coordenada do centro da imagem e as matrizes ao redor de M são as matrizes translação. Isso é feito para que a transformação M tenha o centro da imagem como ponto de simetria.

$$\begin{bmatrix} x^* \\ y^* \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} M \begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

Além disso, ainda é aplicada uma interpolação bilinear como *anti-aliasing* [14, 15]. Durante o treinamento da rede, novas imagens de entrada são geradas a cada época a partir da transformação das imagens originais. A Figura 5 mostra a imagem original, no painel (a), e diversas transformações, nos demais painéis, aplicadas substituindo M da equação (3) por combinações (multiplicação matricial) das transformações da equação (2). Tais transformações não

⁵ <https://www.stsci.edu/~dcotrilog/Intro.html>

⁶ <http://www.image-net.org/>