Product Classification - Training images data cleaning

The following commands need to be run where **pwd = /home/imart/excavator/child\_mcat/train/**  :-

**remove current file extensions(to make all the file names coherent i.e. without any extension):**

for d in ./\*/;do (cd "$d" && for f in \*;do (mv $f $(echo ${f%\*.\*}));done); done

**add .jpg extension (now adding the required file extension):**

for d in ./\*/;do (cd "$d" && for f in \*;do (mv $f `basename $f `.jpg;);done); done

**make image format as .jpg:**

for d in ./\*/;do (cd "$d" && for f in \*;do (mogrify -format jpg \*.jpg;);done); done

**to format all images except for those having JPEG format:**

for d in ./\*/;do (cd "$d" && for f in \*;do (file $f| grep -v JPEG|mogrify -format jpg $f);done); done

**rename files starting with hyphen(-):**

for d in ./\*/;do (cd "$d" && echo $d  && for f in -\*;do ([ -f "$f" ] || continue && mv -- "$f" "${f/-/x}" );done); done

**Also, some cases were found where .jpg.jpg was found, so**

**rename .jpg.jpg extensions to .jpg:**

for d in ./\*/;do (cd "$d" && for f in \*;do (rename 's/.jpg.jpg/.jpg/' \*);done); done

**Procedure to remove images below a particular image resolution:**

* Make a script using vi editor having the below function:

(Here the images having either width or height below 300px are being removed)

Eg. $ vi imageresolution

function rmImgBySize

{

  minX=300

  miny=300

  xres=`identify -format "%[fx:w]" "$1"`

  yres=`identify -format "%[fx:h]" "$1"`

  if [ $xres -lt $minX ] || [ $yres -lt $minY ]

  then

    rm "$1"

  fi

}

* Load the function on command prompt in the same path where above file is kept:

export -f rmImgBySize

* Run the following command to execute the above function iteratively for all the directories in the highlighted part(change it according to your required path):

for d in ~/excavator/traintest/\*/;do (cd "$d" && find . -iname "\*.jpg" -exec bash -c 'rmImgBySize "$0"' "{}" \;); done **>** **/**dev**/**null 2**>&**1