

There were two sets of duplicates between the test and train sets:

Duplicate Pair 1

```
filename species filesize nrow ncol                               md5
1 159128         8      2435   103    46 df7a547fc67ca47b38e83be45adf1c41
```

train\appendicularian_straight



159128.jpg

```
filename species filesize nrow ncol                               md5
1 108193         2435   103    46 df7a547fc67ca47b38e83be45adf1c41
```

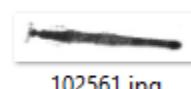


108193.jpg

Duplicate Pair 2

```
filename species filesize nrow ncol                               md5
2 102561         13      3551   176    48 2bc6fe31123d532d690ea93cdfe4c401
```

train\chaetognath_sagitta



102561.jpg

```
filename filesize nrow ncol                               md5
2 89438          3551   176    48 2bc6fe31123d532d690ea93cdfe4c401
```



89438.jpg