

Lightroom Duplicate Image Finder



*Scan your Lightroom image catalogue to
find suspected duplicates.*

Introduction 1

 Contact and Support 1

 Requirements 1

 Legal Stuff..... 1

What's Included 1

 New Commands 1

 New Smart Collections 2

Installation 3

 Windows 3

 Mac..... 3

 Setting up Lightroom..... 4

Searching..... 6

 Filtering your Results..... 7

Finding your Results 8

 Limiting Results 8

 Deleting Unwanted Photos 9

Known Problems 10

 High-speed shooting may result in incorrect duplicates being found 10

INTRODUCTION

This is a straightforward plug-in which will attempt to find duplicate images in your Lightroom catalogue using the EXIF data associated with each photo.

All photos that are potential duplicates are shown. The decision on which ones to keep or discard is yours – this is too dangerous to leave to an automated process! There are easy methods of reducing the amount of work that needs to be done after isolating duplicates using the plugin.

CONTACT AND SUPPORT

You can get support or ask questions in the official helpdesk:

<http://www.lightroom-plugins.com/support/index.php?/Lightroom/Troubleshooter/Step/View/6>

There are several FAQs available at:

<http://www.lightroom-plugins.com/faq.php>

REQUIREMENTS

Lightroom 3 or higher.

LEGAL STUFF

This plug-in is not specifically endorsed by Adobe Inc®.

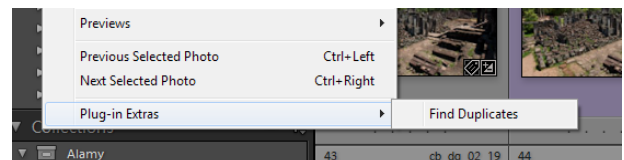
WHAT'S INCLUDED

NEW COMMANDS

One new command has been added to the “Library” menu in Lightroom:

1. *Find Duplicates*

This performs the scan of the catalog, marking any suspected duplicates as it goes.



Please note that there are several different “Plugin Extras” menus in Lightroom. The new commands for this plugin are added to the “Library” menu and nowhere else.

To make it easier to find your images, there are several new Smart Collections provided. Smart Collections automatically find images in your catalog without having to be manually refreshed when something changes. These new Smart Collections are grouped together under a single Set called “Duplicate Photos” and are all created automatically the first time you run the plugin.

- ***All Duplicates***

Shows all images which have been identified as being potentially duplicates. The plugin doesn’t attempt to work out which one you might want to keep so *all* matches are shown in this collection. Please remember that the plugin has no idea of which one of many similar photos you may want to keep.

- ***File - Largest***

This contains, for each set of duplicate photos, the one with the largest filesize. If any of the duplicates are currently offline or missing, none of that set will be included here since it is not possible to get the filesize of a missing file.

- ***File - Smallest***

This contains, for each set of duplicate photos, the one with the smallest filesize. If any of the duplicates are currently offline or missing, none of that set will be included here since it is not possible to get the filesize of a missing file.

- ***Megapixels - Least***

This contains, for each set of duplicate photos, the one with the smallest number of megapixels.

- ***Megapixels - Most***

This contains, for each set of duplicate photos, the one with the highest number of megapixels.

- ***Date - Oldest***

This contains, for each set of duplicate photos, the one with the oldest “Date Time” value. This is different to the “Date Time Original” value and is typically the date on which this copy of the photo was created.

- ***Date - Newest***

This contains, for each set of duplicate photos, the one with the newest “Date Time” value. This is different to the “Date Time Original” value and is typically the date on which this copy of the photo was created.

- ***Edits - Least***

This contains, for each set of duplicate photos, the one with the least number of edits recorded by Lightroom.

- ***Edits - Most***

This contains, for each set of duplicate photos, the one with the highest number of edits recorded by Lightroom.

INSTALLATION

This plug-in adds a new field to your Lightroom database. This is used to store a number indicating which images are seen as duplicates. Each set of duplicates will have a different number.

Before you install this, you should make sure you have a backup of your Lightroom catalog. If you do not already have automatic backups enabled, you can perform a single backup by going to [“Edit→Catalog Settings...”](#) and changing the “Back up catalog” list to “Next time Lightroom starts only”, and restarting Lightroom.

WINDOWS

AUTOMATIC INSTALLATION

The easiest way to install on Windows is to use the MSI installer. Simply double-click and run through the prompts that appear.

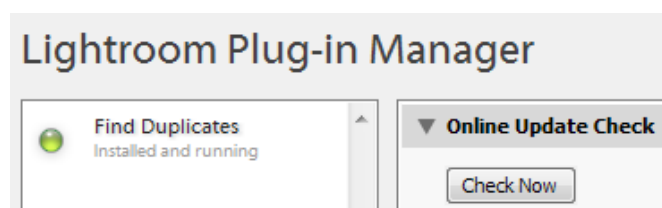
MANUAL INSTALLATION

If you prefer to install manually, download the ZIP file instead of the MSI. When it has downloaded, follow the instructions below for installing on the Mac; the steps are identical on Windows.

MAC

Download the ZIP and extract it to your desktop. You should see a folder called “DuplicatesPlugin”, containing the plug-in and a folder called “Smart Collections”. You can place the plugin anywhere you like. Open Lightroom’s Plugin Manager from the File menu, and click the “Add” button at the bottom of the list. Use the Finder window to select the plugin.

It should now be listed in the Plugin Manager as “Installed and Running”.



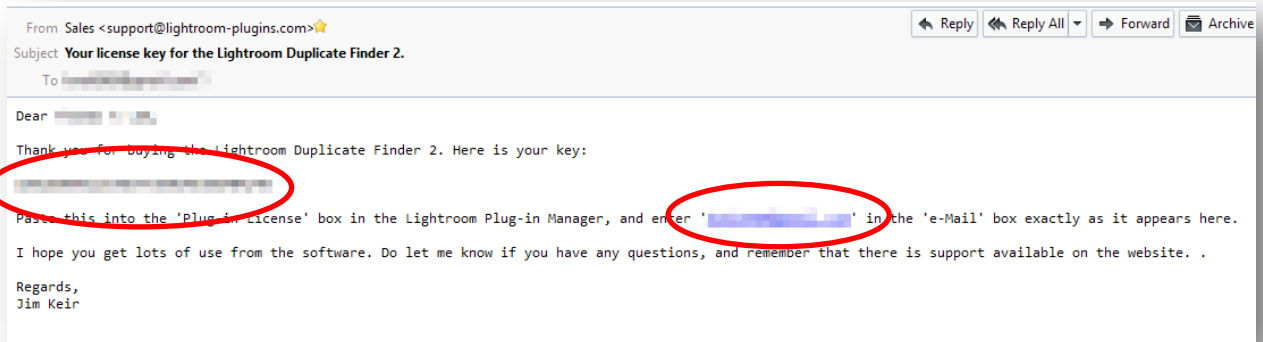
SETTING UP LIGHTROOM

There is nothing more you need to do to prepare Lightroom for using the plugin.

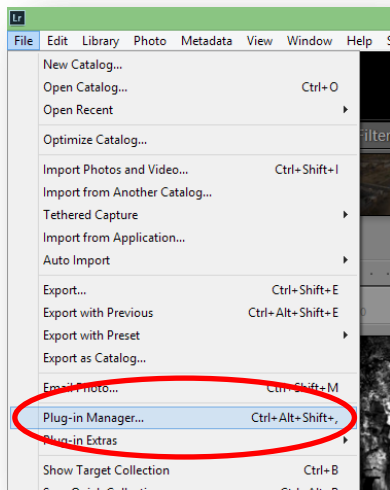
ENTER YOUR LICENCE KEY

If you have bought the plug-in, you should receive a licence key in an email. Without this, the plug-in will stop after finding a small number of duplicates. To enter the licence key, simply copy and paste both the key and your e-mail address from the email to the box in the Plugin Manager when the Find Duplicates plugin is selected.

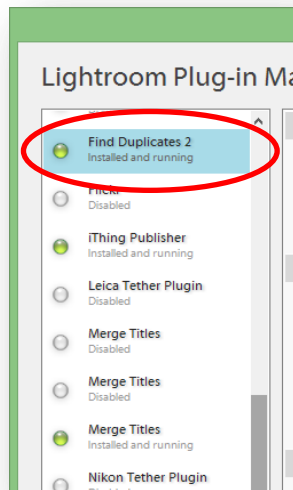
The Plugin Manager can be found on the File menu. If you would like a more detailed tutorial on how to use this, or on other aspects of Lightroom usage, please visit <http://www.learn-to-lightroom.com>.



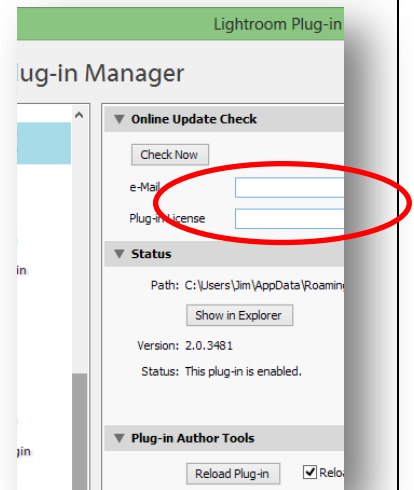
1) Find your license.



2) Open the Plugin Manager.



3) Select the correct plugin.



4) Paste in your email and license key.

SEARCHING

This assumes you already have some images stored in Lightroom.

Select the “[Library→Plug-in Extras→Find Duplicates](#)” menu item. A dialog will appear:

The plugin can search in one of three ways:

- *All photos*

Your entire catalog is searched for duplicates.

- *Within Selected*

Only duplicates within the currently selected photos are found.

- *For Selected*

Duplicates of the selected photos are searched for in the entire catalog.

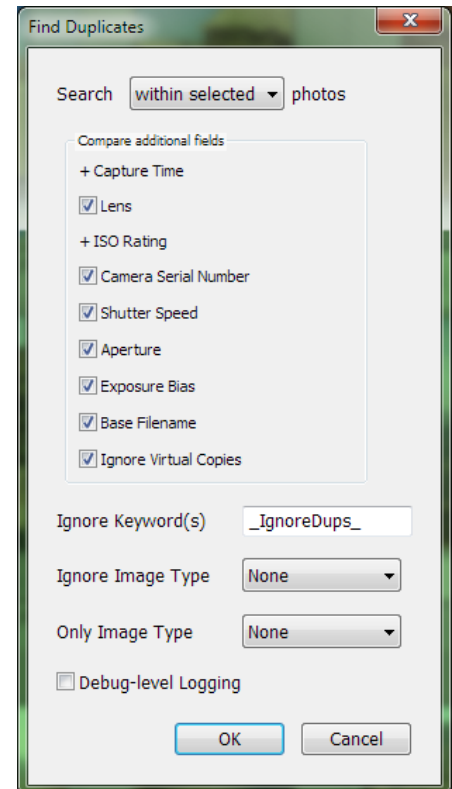
Next you can choose what EXIF data to compare, to determine if one photo is a duplicate of another. Capture Time and ISO Rating are always used while the others are optional, although they are likely to be always switched on. You could use these to decide whether to consider bracketed shots to be duplicates or not, for example.

The exception is the final option, “Base Filename”. All the other items are attributes of the photo itself and so are good candidates for deciding whether two photos are duplicates. The filename could easily change though, so you will need to decide whether to match filenames or not.

A limitation of EXIF data is that the timestamp is limited to one-second accuracy. This means that photos taken very close together with no settings changes can be picked up as duplicates. If you are bracketing then usually the exposure or aperture will change, so this can be used to filter these out. Photos taken this close together will usually appear to be very similar and are still probably worth seeing as duplicates, but if you want to keep a high-speed sequence like this then you can mark them as “not duplicates” permanently using the “Not a Duplicate” tickbox in the Library module’s metadata panel.

The other options all affect which photos are included in the results, allowing you to tailor the search to your own requirements. When ticked, the options like “Aperture” and “Exposure Bias” require those values to be matched before two photos are considered duplicates.

Virtual copies are handled slightly differently. With “Ignore Virtual Copies” ticked, no virtual copies will be included in the results in most cases. The exception to this is where you have chosen to search “Within Selected” or “For Selected” photos, and you have selected a virtual copy. In this case, since you have explicitly chosen a virtual copy to work with, it *will* be included in the search results if it has duplicates.

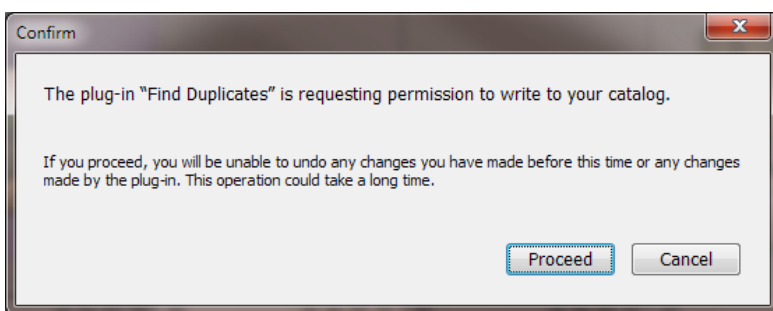


FILTERING YOUR RESULTS

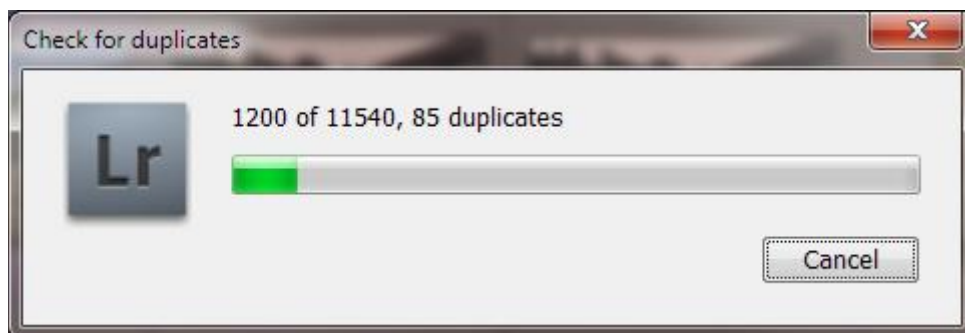
You can choose to ignore photos in several ways. One way is to have a specific keyword (or list of keywords) and to tell the plugin to ignore any photos with this keyword. For example, if you have uploaded processed copies of certain photos to an online picture agency, you can ignore all these processed copies by adding a non-exportable keyword to them and entering that keyword here.

Another way is to use the “Duplicate Search” option which can be found in the Library’s Metadata panel under “All Plugin Metadata”. By setting this to “Ignore”, the duplicate finder will never consider this image to be a duplicate.

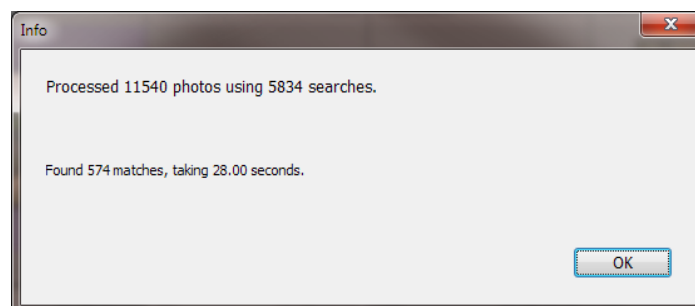
To begin the search, click “OK”. A warning dialog will appear:



The only change to be written to the catalog is the unique identifier for each set of duplicates. Click “Proceed” to begin the search itself. A progress window will appear:



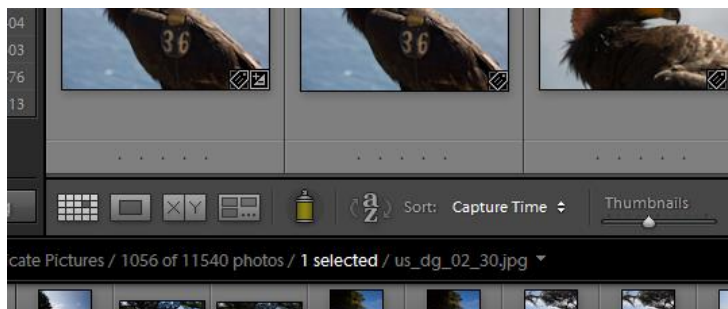
Once the search has completed, a summary is shown:



FINDING YOUR RESULTS

Once a search has completed, you can view the results. The search has filled in a new field in the database called “Duplicate Id”, and it is this that the Smart Collection searches on. Any photo which is considered to be a duplicate will have a number in this field, the same number being assigned to each duplicate of that photo.

To view the results, simply select the “Duplicate Pictures 2” Smart Collection. Unfortunately it's not possible to sort the results by the “Duplicate Id” field, but the next best thing is to sort by “Capture Time”:



Since the capture time will be identical only for duplicate images, this works just as well. It's now up to you to decide what to do with your duplicates. You can, of course, use Lightroom's normal Library Filter tools to limit the selection, for example by file extension, so that you can delete unneeded photos.

LIMITING RESULTS

With the “Found Duplicates” Smart Collection displayed, you can apply any combination of normal Lightroom filters to help you isolate the copies you want to keep from those you want to get rid of. The plugin deliberately makes no attempt to make this decision for you! From the online FAQ:

The plugin will identify all images that are likely to be copies of each other, based on the image metadata. Many people ask why it doesn't automatically delete duplicates, or why it doesn't exclude the original.

This is simply down to safety. For a commercial photographer, losing a valued original image may be disastrous. Would you really want a bit of software deciding which photo to delete based on some kind of pre-written rule?

Suggestions have included:

- *Keep the one with the shortest name, assuming that duplicates will have a number or "copy of" in the name. So, we delete "Photo (Original).jpg" and keep "Photo (Web).jpg". Wrong.*

- *Delete JPEGs, keep the RAW. But what if you shoot in JPEG? What if you convert to DNG and use that as the master, and the RAW is the one you would want to get rid of?*
- *Keep the oldest one, on the assumption that the oldest is the master. So, we delete the one to which you have made loads of production edits and keep an early pre-view you exported for your client. Wrong.*
- *Keep the newest one, on the assumption that it will be the one with the most work done to it. So, we delete the older original and keep the web-sized, watermarked export you prepared for a customer. Wrong.*
- *Keep the largest one, on the assumption that all smaller ones are previews. So, we delete an original scanned copy and keep a resized-for-printing copy with a banner across it. Wrong.*

*In short, no matter what rules you have for identifying which image is the "original" and which are "duplicates", there will be someone else who does the opposite. The plugin will help you to **identify** which images to look at by eliminating most of your library, and letting you display only the potential duplicates next to each other. You can then use Lightroom's own tools to identify, mark and delete any that you don't want, according to your own rules.*

In short, every person will define "original" as something different. Rather than take the risk of automatically identifying the wrong copy of a photo for deletion, the plugin leaves this up to you.

For a filtering tutorial, visit <http://www.learn-to-lightroom.com> .

DELETING UNWANTED PHOTOS

Once you have limited the photos in your view, it is possible to delete unwanted photos. If you try and delete from the Smart Collection, Lightroom will warn you that this is not possible. It assumes that you are trying to remove a photo from the collection, and since this is a Smart Collection which is populated automatically, this is not possible.

Removing photos is still easy though. You need to use Lightroom's "Reject" feature. With this, you can reject photos whether they are part of a collection or not. The easiest way is to press X on the keyboard, but as usual with Lightroom there are many ways of using the same feature. Once you have rejected some photos, you can finally delete them all at once using the "Delete Rejected Photos" command on the "Photo" menu.

HIGH-SPEED SHOOTING MAY RESULT IN INCORRECT DUPLICATES BEING FOUND

As mentioned above, it is a known limitation of EXIF data that timestamps are accurate only to the nearest second, even though the camera stores much more accurate information in the original RAW file. Since this plugin works with EXIF data, this means that high-speed sequences where no bracketing is done will be identified as duplicates.

In practise these shots are likely to be very similar and it may be useful to have them highlighted anyway. Unfortunately without adding the ability to read the original RAW files to the plugin, there is no way of getting more accurate timestamp information out of Lightroom.

Once identified, high-speed sequences like this can be permanently marked as “not duplicates” in the Library’s Detail panel on the right-hand side of the screen.