



PHOTO FORENSICS

USER-GUIDE

Table of Contents

Welcome to Photo Forensics	3
The Big Duplicate Problem?	3
So What?	3
How Does Photo Forensics Help?	4
Who is Photo Forensics aimed at?	4
A First Timers Guide To Removing Duplicates.....	5
Where Do I Start?	5
Then what?	6
The User Guide.....	7
The Main Window	7
Starting A Scan	10
The Scanning Process.....	13
Scan Summary	15
Duplicates Preview	16
Removing the Duplicates.....	19
Settings	21
Frequently Asked Questions.....	23
Why Did Photo Forensics Miss Some Of My Duplicates?.....	23
What's The Difference Between Duplicates and Similar?	23
How Can Similar Be Detected By A Computer?	23
What Photo File Formats Are Supported?.....	23
IT's Enumerating Files Twice – Are You Mad?	24
Why Is This 64-Bit Only?	24
Final Thoughts	25

WELCOME TO PHOTO FORENSICS

Automatically finds similar and exact duplicates of photos including the Apple HEIC format. Using an advanced forensic and rules engine, duplicates are discovered, originals selected, and presented to you for preview before committing to any duplicate removals.

If you're new to removing duplicates, then read on, but if you're a seasoned de-duplicator, you may want to jump right into the user-guide [here](#)

THE BIG DUPLICATE PROBLEM?

As we take more and more pictures, and share them with friends and family, our library of photos increases, and because they are precious to us, we back them up (you do backup your photos, right?). iCloud and Google Cloud are obvious places to store them (nicely built into our phones) but that doesn't protect them, it's simply a place to store them, and of course being in the cloud makes them vulnerable to being stolen. So, many people keep a copy on their computer, or (if you're tech savvy) on a home storage solution like Network storage, USB drives, or even on other computers. As we keep more and more copies, we don't sift through them to make sure we haven't copied them before, and so, we end up with duplicates, everywhere.

There's also another reason we end up with duplicates, as photos are so easy to take, we take lots, of the same thing, just to make sure we got the right picture, and then forget to remove the ones we don't want. These photos are not identical duplicates, they are similar duplicates.

SO WHAT?

A good question. Well, you have to store them somewhere, and that costs money, your money. If they're in the cloud, you are paying a subscription; if they're on a disk, you had to buy the disk. When the storage fills up or you run out of space for other stuff, you have to buy more.

There is also the issue of searching photos, the more you have the harder it is; the more duplicates or similar photos you have, well that makes searching a whole new level of difficult.

HOW DOES PHOTO FORENSICS HELP?

Photo Forensics analyses your photos and finds all those duplicates, it also finds the photos that look the same, for example, taken in quick succession, or have been edited to look brighter, etc. It even finds similar photos that have been converted to different format, which is very handy if you have an iPhone and converted Apple's HEIC format to the more common JPG format.

WHO IS PHOTO FORENSICS AIMED AT?

Everyone, from individuals and family members, to professional photographers, marketing departments, or just a collector of imagery. Anyone with a large collection of photos knows the problems of having duplicates in different folders.

A FIRST TIMERS GUIDE TO REMOVING DUPLICATES

This section is designed to guide you through the process of removing duplicates and similars. This is not a technology guide to the Photo Forensics app, that's further down, but a "how do I approach this" and successfully remove them.

WHERE DO I START?

We're going to assume you have a lot of duplicates; if you only have a handful then it's actually pretty easy – run Photo Forensics and press the 'remove duplicates' button.

A lot of duplicates basically means 'more than you're prepared to sit down and make decisions about' – if you're happy going through thousands of duplicates, skip to the user-guide.

So, we start by knowing where all your photos are, if they're in the cloud, you'll have to download/sync them, if they're on a network, local disk, scattered about your disks – that's not a problem.

Photo Forensic gives you several 'scanning' options. The reason is to help you manage the duplicates you're going to find.

We recommend you start by using the "Intelligent Scan". This is the fastest scan type and looks for photos that are identical. It checks their file size and a sample of the photo and its metadata (information about the photo put in by the camera/phone, for example: Date Taken). This will give you a good idea of the size of problem you have. Note that this is looking for exact duplicates, photos that are 100% identical, in size, colors, contrast, metadata, etc. All the scan types ignore filenames, so renamed photos will still be detected.

It is possible (but unlikely) that this scan will give some false duplicates; typically when advanced editing tools don't update the metadata correctly. If you think that might be an issue, run the "File Scan" instead, this analysis compares the whole file, it's much slower but 100% accurate.

Now you have a list of duplicates you need to decide if the automatic selection by Photo Forensics is what you would choose or not. If it is, you can click the “remove duplicates” and they’ll be sent to your recycle bin (you only get your disk space back when you empty the recycle bin).

What if you want to decide which are duplicates and which are original yourself? And you have 10,000 duplicates? – Our recommendation (until we implement a new feature for this exact scenario, and it is coming) is to cancel and start again using less folders to search, or sub folders, gradually adding more folders as you get through each scan. This is why this particular scan mode is important, it allows you to scan very quickly, remove the duplicates and start again.

THEN WHAT?

Once you have no identical duplicates, use the “similar” scan type. This will find all your photos that look the same, they’re not identical, but close enough for a human to go “they look identical” or “they’re part of a series of shots with small changes” or “oh yes, I brightened that picture up”.

This type of scan is slow. We’re decoding each photo and turning it into a machine learnable fingerprint – basically so the computer can “see” it like a human would. This is very complicated and requires a lot of processing power, until recently this kind of technology was only available in the cloud with machine learning systems. It takes time, but the results are amazing.

Now that you have your Similar, you remove them the same way as before; if you have a lot, then try scanning less folders or sub-folders as before.

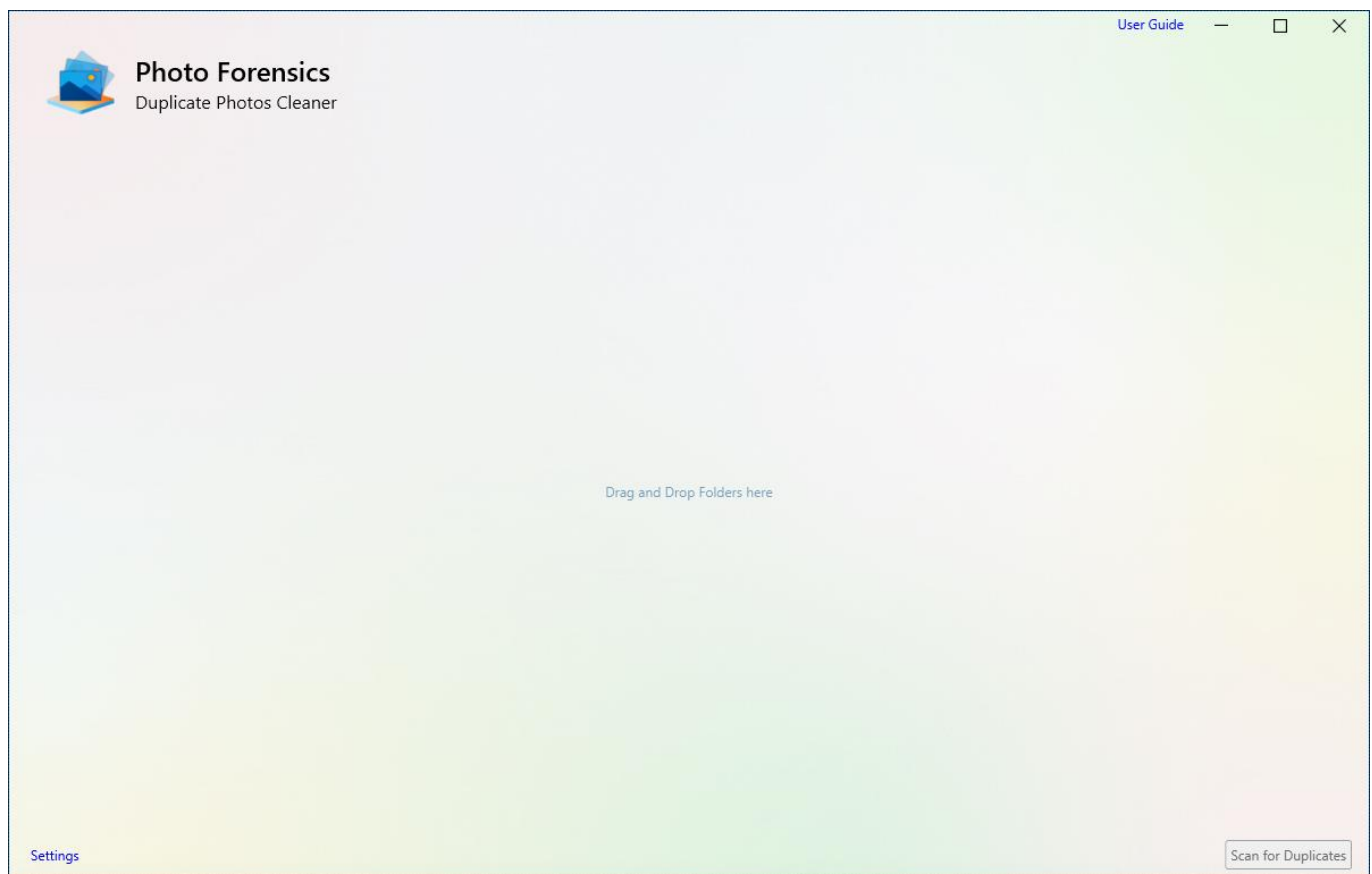
THE USER GUIDE

This section shows you what each part of the product does and the general “flow” of removing duplicates.

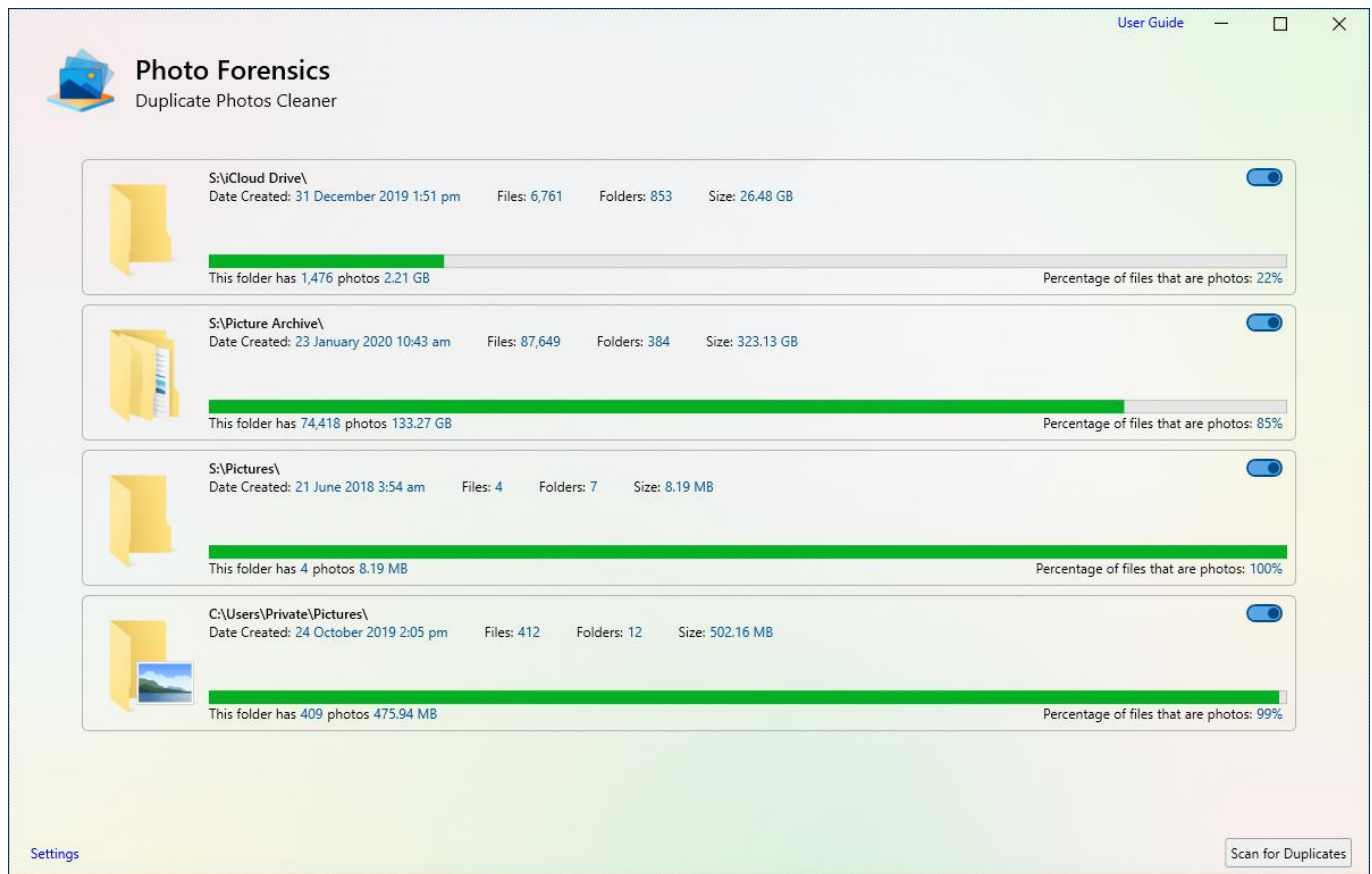
Photo Forensics is designed to be as easy as possible, hiding the complexity of what it’s doing behind the scenes, and using best practice settings to find your duplicates and similars.

THE MAIN WINDOW

Photo Forensics starts with a blank canvas. Simply drag and drop the folders or drives from Windows Explorer that you want scanning.

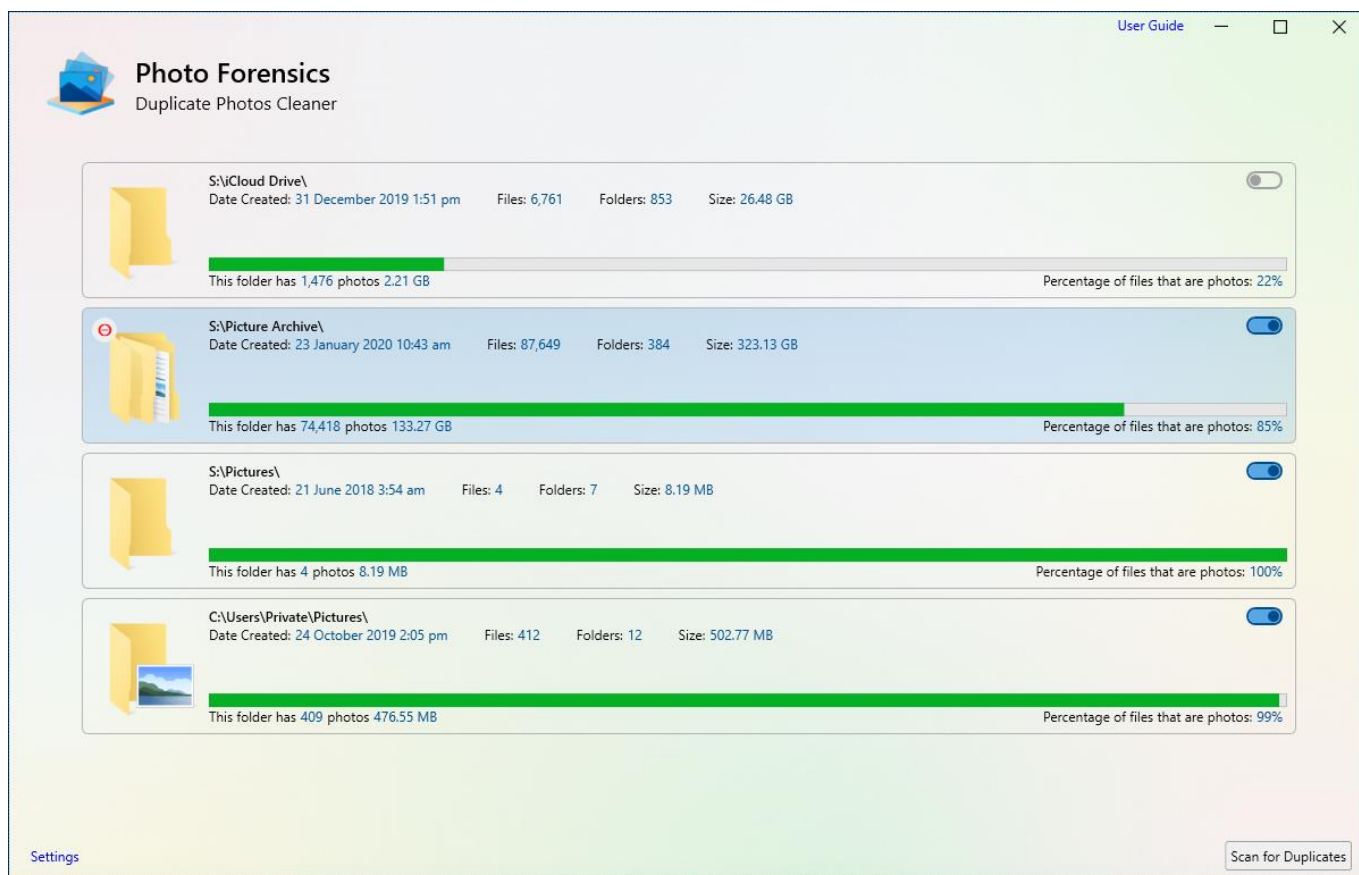


You can drop as many folders as you want, you can also drop drives, like “Drive C”, USB drives, or network drives.



When you drop a folder onto Photo Forensics it automatically starts examining it, looking for photos, don't worry, this is for information only and so you can see the ratio of photos in the folder (and sub-folders). You don't have to wait for it to complete before starting a scan.

Photo Forensics remembers (default option) the folders and drives you drop into the window and will auto-load them each time you start a new session.



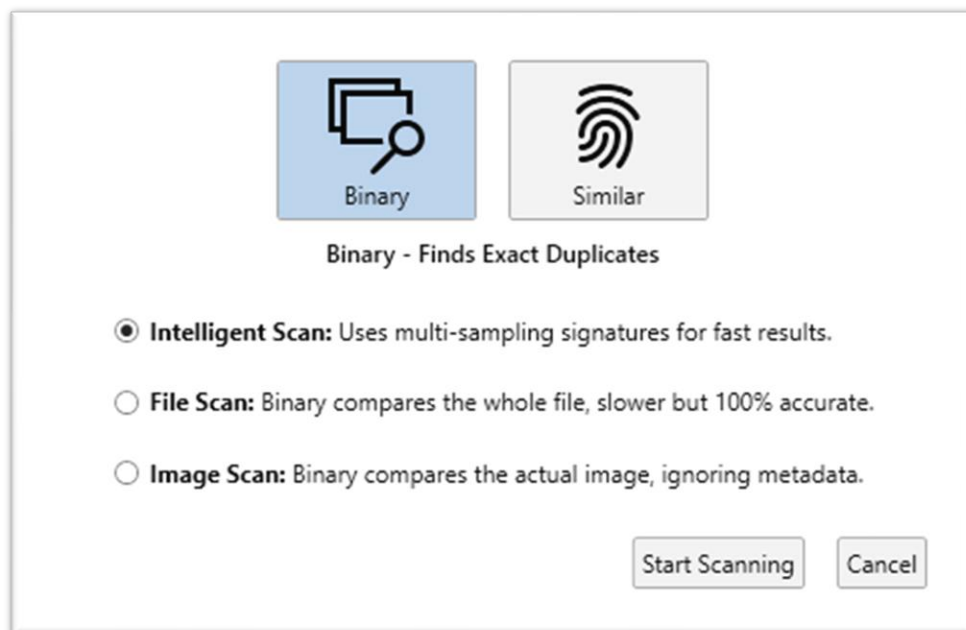
In the screenshot above you can also see that hovering the mouse over an item, highlights it and shows a 'remove' button in the top left of the item. Clicking the 'remove' button will remove this item from the list.

You can also see a 'toggle button' on the top right of each item, this determines if the item is part of the scan. Turning it off will remove that folder from the scan, analysis, and duplicate preview. This is useful when you run regular scans and don't want to keep removing and adding folders to the list, you simply disable them instead.

STARTING A SCAN

Once you have at least one item in the main window, the “Scan for Duplicates” button is enabled, and you can start a scan. A scan simply examines the photos in the folders you’ve dropped, looking for duplicates, don’t worry, nothing is being changed and you can cancel it at any time.

When you press the “Scan for Duplicates” button, a popup will prompt you to choose the type of scan you want. There are two main types of scan, Binary, and Similar.



Binary scans are looking for exact duplicates. These are copies of originals that typically come from downloading photos twice (easy to do from iCloud or Google), or from copying folders around, especially if you have multiple backups.

Photo Forensics has three scan modes for binary comparisons:

1. **Intelligent Scan**

Uses multi-sampling signatures for fast results. This is the fastest scan mode, it doesn't decode the photo or check the whole file, so it is can read little bits of each file very fast. It is very quick and reliable.

2. **File Scan**

Binary compares the whole file, slower but 100% accurate. This is slower than an Intelligent scan as it reads the whole photo file, however, it doesn't decode each image, so it will run as fast as your disk will let it.

3. **Image Scan**

Binary compares the actual image, ignoring metadata. This is the slowest of the binary scans as it has to open each photo and decode it into a picture, the picture is then given a cryptographic signature for comparisons. This can go as fast as your disk and processor allows, although it will use your graphics card to speed up the decoding where available.

A sub note on HEIC files (Apple's native photo format), these files are amazingly complex to decode and will be slow to analyze. Apple uses hardware decoding and encoding to do this which is not available on Windows, the format is also a patent nightmare (7,000+ patents) and this hinders the development on the Windows platform, fortunately, Photo Forensics has built-in support for HEIC files.

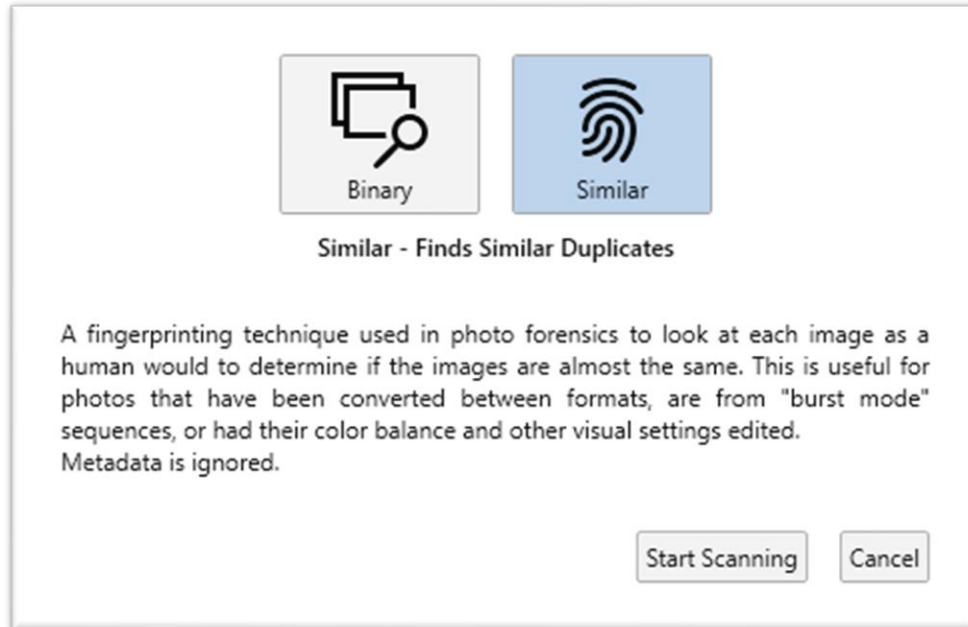


Photo Forensics can also scan for similars.

1. **Similar Scan**

A fingerprinting technique used to look at each image as a human would to determine if the images are almost the same. This is useful for photos that have been converted between formats, are from "burst mode" sequences, or had their color balance and other visual settings edited. Metadata is ignored.

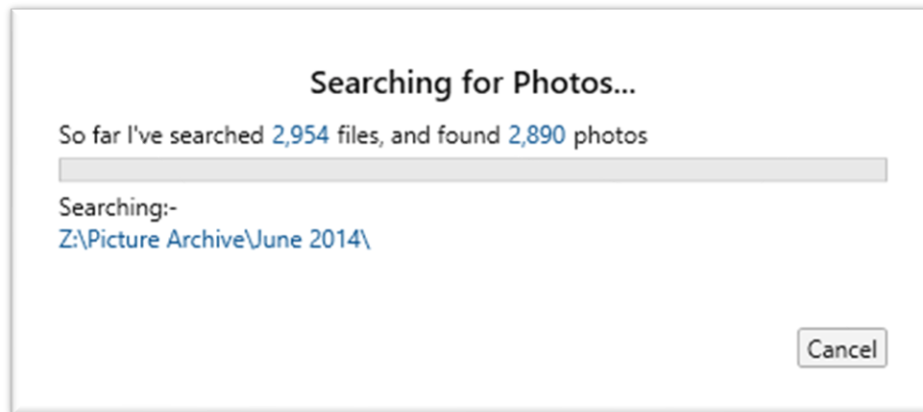
This is the slowest of all scans due to the processing required to 'see' photos like a human would.

Once you have selected a scan method, click the 'Start Scanning' button to start the analysis.

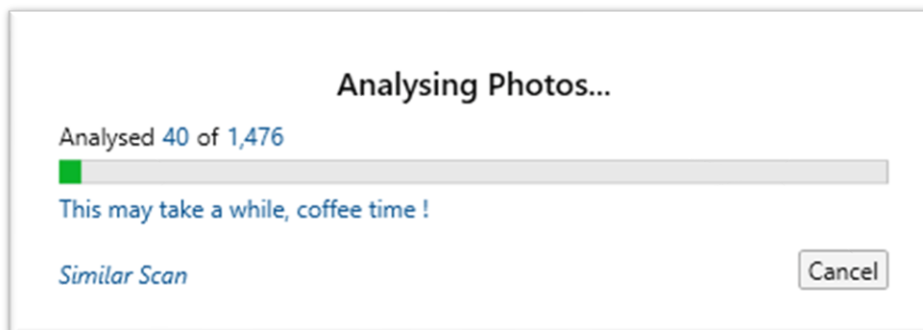
THE SCANNING PROCESS

There are several parts to find you duplicate photos which Photo Forensics will show you what's happening along the way.

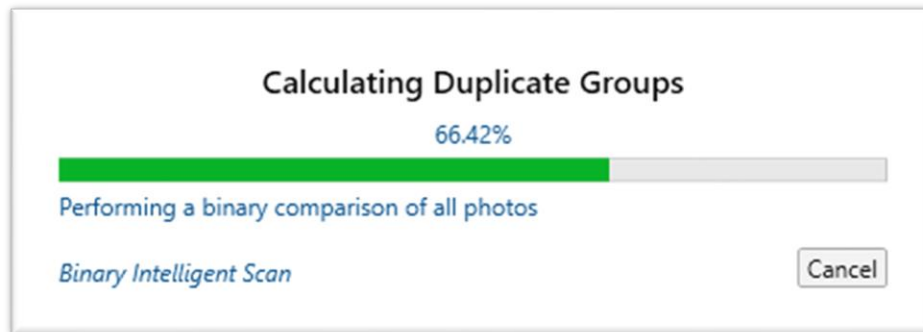
Step 1 is to find all your photo files. This is very quick, and you may not even see the popup showing the progress. If looking for your photos is taking some time, you'll see the popup below: "Searching for Photos...", this can happen when there are a lot of files to find before the analysis starts, or on slow media such as network drives.



Step 2 analyses the photos; this is the time-consuming part.

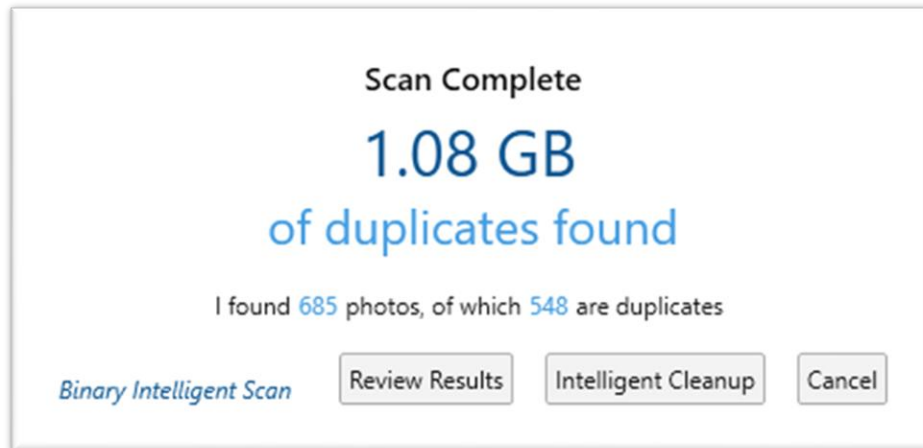


Step 3 uses all the information from Step 2 to group the duplicates, this is Photo Forensics seeing like a human (for similars). This is very quick, but it can take a little while for very large archives.



SCAN SUMMARY

Once the scan has completed, a summary popup shows you what was found:



From here, you have several options on how you'd like to move forward:

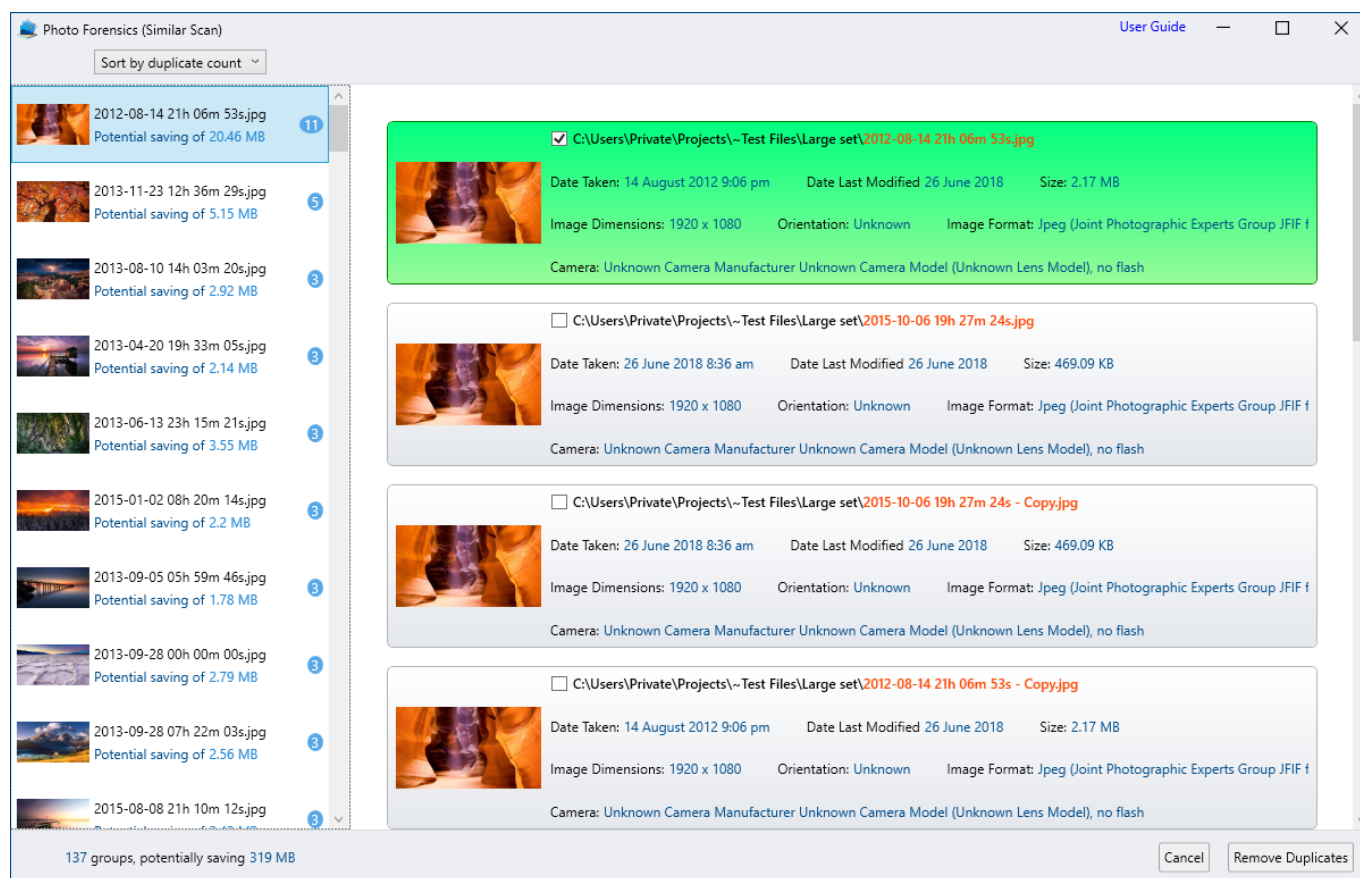
Cancel: This takes you back to the main window, maybe you'd like to run a different type of scan (for example, if no duplicates were found) such as a similar scan.

Intelligent Cleanup: This will open the "Removal Policy" window, so don't worry, nothing will happen just yet. This is shown further down in this guide.

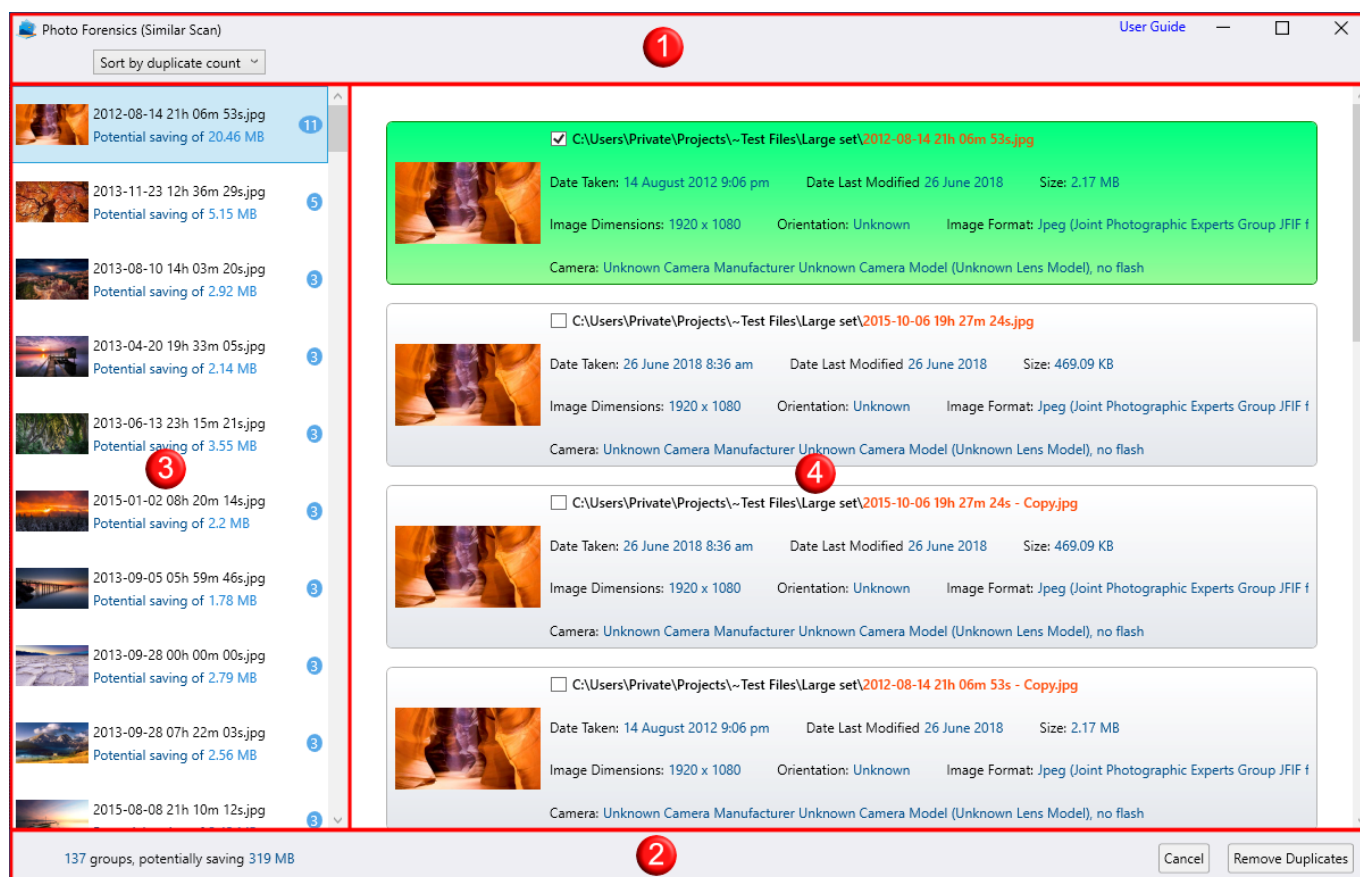
Review Results: This allows you to review the results, the groups of duplicates, and which photos the rules engine determined were the originals and which are duplicates. This is discussed in the next section.

DUPLICATES PREVIEW

Once the analysis has completed, a preview window opens to allow you to confirm and manually override the selection process. Photo Forensics uses a sophisticated rules engine to determine which photos are original (in green and ticked) and which are duplicated (not green or ticked).



Let's take a look at the preview screens sections in more detail.



1. The Toolbar

This section has one dropdown box for sorting the group list (section 3) by the number of duplicates, or by the size in savings. The window title also reminds you of the scan type you selected.

2. The Footer

This section shows the number of groups in the group list (section 3) and the potential savings if all the duplicates were removed.

The “Cancel” button will cancel this session and return you to the Main Window, basically starting again from the start.

The “Remove Duplicates” button will start the removal of duplicates, that is, any file not selected (green and checked).

3. Group List

This section groups all the duplicates. Each group has a thumbnail showing the photo that has duplicates; it also has a number showing the number of duplicates in the group, so a “1” means there is one photo (that’s 2 files, one original, one the duplicate).

Each group also shows the filename associated with the group and the saving that would be made should all the duplicates be removed.

Selecting a group update section 4, the duplicates preview list.

4. Duplicates Preview List

This section shows the photos that Photo Forensics has determined to be identical or similar (depending on the file scan type you chose earlier).

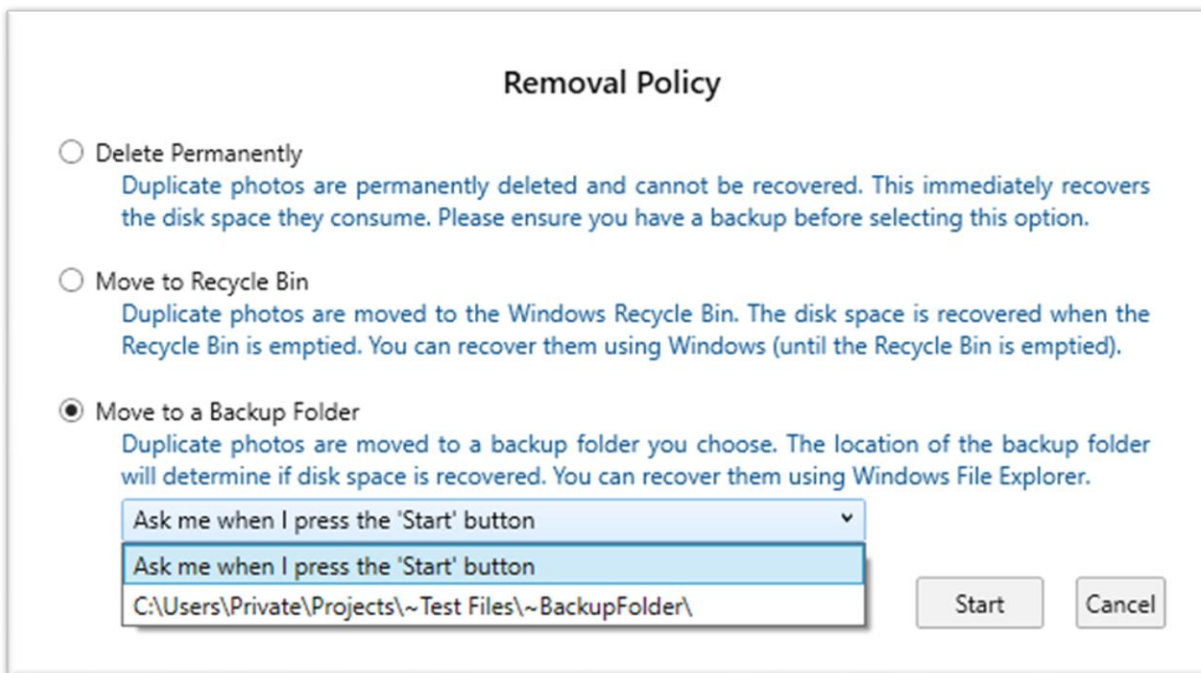
Photo Forensics uses a sophisticated rules engine to determine which photo in the group is the original and automatically selects it. This turns it green and checks the box next to the filename.

You can unselect it and select other photos in the group to be the original. You can also select more than one photo as original.

Photos that are selected (green) will not be removed from your folders.

REMOVING THE DUPLICATES

Once you've selected "Remove Duplicates" or "Intelligent Cleanup" a new window opens with the "Removal Policy", don't worry, nothing happens just yet.



The removal policy determines what happens to the duplicate photos, they're self-explanatory, however, for clarity, they can be:

Deleted Permanently: Which will remove them forever.

Moved to the Recycle Bin: Which will move them to Windows Recycle bin. From here, they follow Windows delete policy which you've set (or left at default) in Windows Storage Sense. To change those settings, go to windows search (or Windows Settings) and search for "Change how Storage Sense frees up space"

Move to Backup Folder: Which will move them to a folder of your choosing. This allows you to back them up in a folder structure that matches their original location, making it easy to recover them or see where they came from in Windows File Explorer.

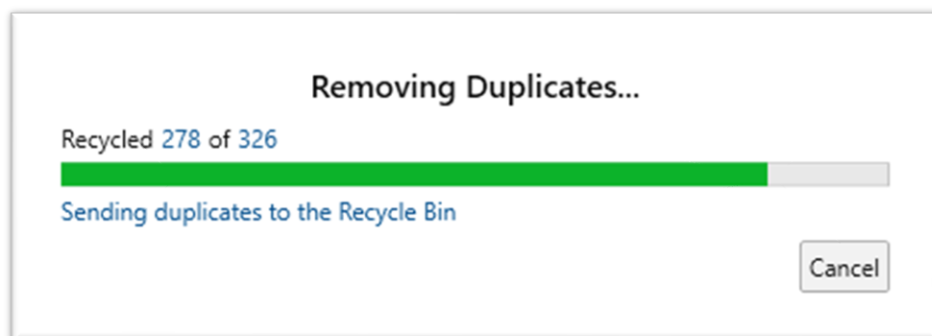
You can select another drive, which would free up disk space where the duplicates came from, or the same drive, which will not free up any disk space until you manually delete them.

Once you select a backup folder, it is added to the dropdown list and made the default for next time, however, you can add more backup folders to the list by reselecting “Ask me when I click the Start button”.

Pro tip – if you want to remove the list of folders in the drop down, hold down the ‘shift’ key when opening this window. That’s either from the summary window, or the preview window, just hold down the shift key when pressing “Remove Duplicates” or “Intelligent Cleanup”.

It’s time to press the ‘Start’ button.

At this point we have to put in the obligatory warning. This action will send the duplicates to the recycle bin (or wherever you chose). You may want to ensure you have a backup before proceeding.

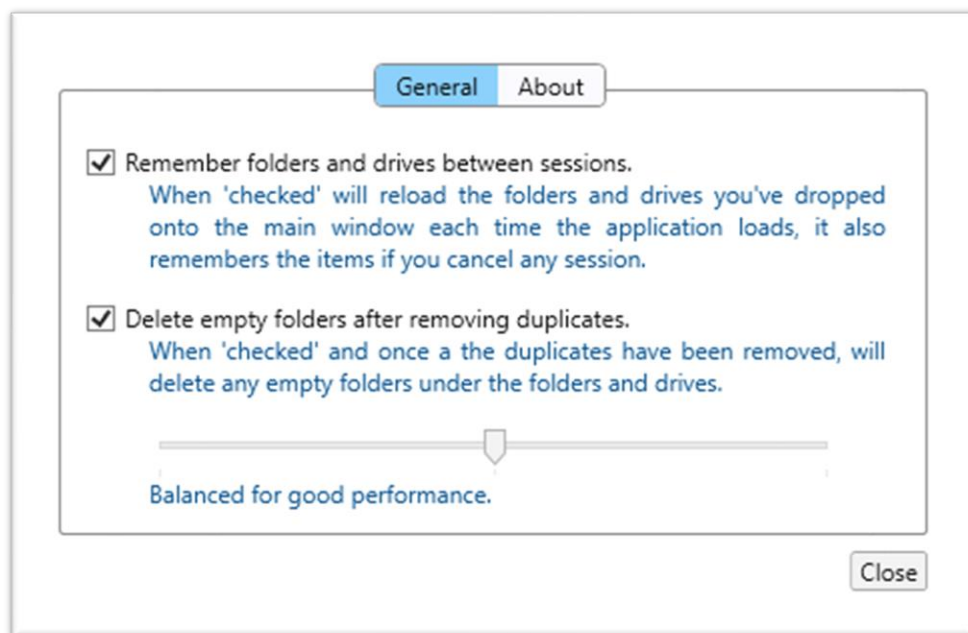


You can of course cancel at any time.

Once complete, any folders that are now empty, because of the removal of duplicates, will be removed (default), then you’ll be returned to the Main Window, duplicate free.

SETTINGS

At the bottom left corner of the main window is a 'settings' button, this will open the settings dialog window like this: -



The settings are remembered between sessions and act immediately. In the about tab, there are buttons for requesting support or rating the app. Most options are explained within the window; however, the performance slider may need a little more detail.

Performance Slider: As mentioned previously, analyzing photos so that they can be viewed similar to a human is an intensive process and by default, Photo Forensics will balance the use of system resources with performance. You can change this behavior by sliding to the left, which is battery friendly at the cost of slower performance, or slide to the right, which gives maximum performance and uses all the resources available.

For example, if the slider is set to battery friendly, only one processor is used, one file is analyzed at a time, and memory usage is kept low. If the slider is to the right, all your processors are used (up to 256), multiple files are analyzed at the same time (up to the performance of your disk or network drive), and memory usage rises (up to your maximum memory).

Does this really make a difference? Yes, typically, having the slider to the right is 4 to 8 times faster than having the slider to the left. This will depend on your computer's capabilities, for example, a computer with 8 processor cores, 8 GB RAM, and an SSD drive will perform a scan much faster than a computer with 2 processor cores, 2 GB RAM and a mechanical disk drive. However, even in the fastest systems, performance will be limited by the slowest component, for example, a computer with 8 processor cores, 8 GB RAM, and scanning a network drive will be limited by the speed of the network and the speed of the network disks.

FREQUENTLY ASKED QUESTIONS

WHY DID PHOTO FORENSICS MISS SOME OF MY DUPLICATES?

It didn't. The exact duplicate analysis is 100% accurate, however, a human 'perceives' what a duplicate is, whereas, a computer examines the parts a human cannot see. See the question on "What's the difference between duplicates and similars"

WHAT'S THE DIFFERENCE BETWEEN DUPLICATES AND SIMILARS?

Exact duplicates are exactly that, 100%, bit for bit, copies of each other. If just one pixel, or one bit of metadata is different, they are not the same. It is common for a photo to be opened by an editor, no changes are made and then it's closed; this in fact changes the photo, all editors 'stamp' their mark on the photo file (in metadata) and while a human may not see it, a computer can.

Similars are photos that look the same to a human. Their metadata may be different, the photo size may be different, the brightness and colors may be different, the photo may even be different but part of a series of 'shots' that have a slight change in each. All these changes don't worry humans, we see past them, or recognize they're almost the same (and wish that cat hadn't walked across the scene when taking them, hence lots of similar pictures).

HOW CAN SIMILARS BE DETECTED BY A COMPUTER?

This is very difficult for a computer. We use a combination of complex mathematical models previously reserved for big servers, but with modern 64-bit hardware and Microsoft's latest technologies, that capability can be brought to the desktop. It is a very processor intensive operation to take a photo and convert it into something a computer can 'see' like a human and then make a 'judgement' on two photos looking similar, so it may take some time, but the magic is worth it.

WHAT PHOTO FILE FORMATS ARE SUPPORTED?

Native, built-in support for HEIC, JPG, JPEG, GIF, PNG, TIFF, and TIFF.

IT'S ENUMERATING FILES TWICE – ARE YOU MAD?

When folders are dropped onto the main window, they are enumerated to find the ratio of files vs photos; the folders also have to be enumerated just before a 'scan' to identify the files that need analyzing.

Don't worry, Photo Forensics is very intelligent, folders are only ever enumerated once. This is done to not only provide you with visual information on your files vs photos, but also to start the enumeration process before you even press the scan button or make your scanning decisions; this improves the performance of the scan and analysis, as that work has been performed up-front.

WHY IS THIS 64-BIT ONLY?

Bringing the ability to 'see' photos like a human is very complex and requires a lot of processing power, but it can be made faster by using more memory (if available). To access that memory requires 64-bit computing; there are other 'technical' reasons for using 64-bit, but it would take more than an FAQ to list them. Our priority is for a stable system; however, we are experimenting with new techniques to increase performance, these will use more and more of the 64-bit core of Windows and it's memory space. As we bring these to market, we will make them optional, give the ability to switch them off, or allow them to be throttled back.

FINAL THOUGHTS

Our aim at Duckhead Software is to make our software as easy as possible to use but allow customization for those people who are more experienced. We constantly develop our software and our priorities go to the feedback we receive.

To that end, we would be grateful if you'd rate our app send us an email with new features you'd like.

Support@duckheadsoftware.com

Team Duckheads.